

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

Кваліфікаційна наукова
праця на правах рукопису

Братусь Олена Вікторівна

УДК 338.27, 681.5.015

ДИСЕРТАЦІЯ
МЕТОДИ ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ
НА ОСНОВІ ДВОБІЧНОГО ЕКСПОНЕНЦІЙНОГО ЗГЛАДЖУВАННЯ
ТА ОПТИМАЛЬНОЇ ФІЛЬТРАЦІЇ

01.05.04 – системний аналіз і теорія оптимальних рішень

Технічні науки

Подається на здобуття наукового ступеня кандидата технічних наук
Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ О.В. Братусь

Науковий керівник Бідюк Петро Іванович, доктор технічних наук, професор

Київ – 2019

АНОТАЦІЯ

Братусь О.В. Методи прогнозування нестационарних часових рядів на основі двобічного експоненційного згладжування та оптимальної фільтрації. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.04 – системний аналіз і теорія оптимальних рішень (технічні науки). – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Міністерство освіти і науки України. Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Міністерство освіти і науки України, Київ, 2019.

Дисертаційна робота присвячена розробці та застосуванню прикладної наукової методології системного аналізу для розв’язання актуальних задач відновлення пропущених значень часових рядів, відновлення істинних закономірностей розвитку досліджуваних процесів, прогнозування часових рядів та виконання автоматизованої обробки часових рядів за допомогою створеної системи підтримки прийняття рішень.

Важливими задачами на сьогоднішній день є відновлення пропущених значень та істинних закономірностей часових рядів, тому що це дає змогу правильно виконати попередню обробку та подальше прогнозування часових рядів. Отримання точніших прогнозних оцінок значень часових рядів є важливим для тих, хто займається прогнозуванням часових рядів у своїй повсякденній роботі. Важливою задачею є розробка нових систем підтримки прийняття рішень для аналізу та прогнозування часових рядів на основі сучасних принципів системного аналізу, оскільки такі вдосконалені системи дадуть змогу виконувати швидко та більш точну обробку даних.

У роботі розроблено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев’язок субоптимального фільтра з пам’яттю два, їх

порівняно з існуючими методами. Обрано кращі методи серед розроблених та застосовано їх для розробки адаптивних фільтрів Калмана. Під час прогнозування середньодобових цін Лондонської біржі металів на цинк розроблені адаптивні фільтри Калмана показали перевагу за значеннями прогнозних характеристик перед методами експоненційного згладжування, авторегресії, авторегресії з ковзним середнім та групового урахування аргументів.

Розроблено метод оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду незалежних псевдовимірів цього параметру, та створено алгоритми прогнозування на основі фільтра Калмана з використанням розробленого методу. Під час прогнозування середньодобових цін на свинець за даними Лондонської біржі металів створені алгоритми прогнозування на основі фільтра Калмана показали перевагу перед методами авторегресії, авторегресії з ковзним середнім та групового урахування аргументів за значеннями прогнозних характеристик.

Розроблено метод двобічного експоненційного згладжування та алгоритм для відновлення пропущених значень часового ряду з використанням цього методу. Розроблено інтегральний критерій адекватності моделі. Описано застосування створеного методу для відновлення істинних закономірностей та прогнозування часових рядів. Під час відновлення пропущених середньодобових значень цін на цинк за даними Лондонської біржі металів метод двобічного експоненційного згладжування за значеннями статистичних характеристик показав перевагу перед середнім арифметичним, авторегресійним підходом та методом експоненційного згладжування.

Розроблено метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. При відновленні значень взаємозалежних показників сталого розвитку за даними Світового центру даних цей метод показав перевагу за значеннями

статистичних характеристик перед методами відновлення середніми арифметичними та експоненційним згладжуванням. Розроблено метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Цей метод показав перевагу перед методом експоненційного згладжування за значеннями прогнозних характеристик при прогнозуванні показників сталого розвитку.

Розроблено метод ковзного двобічного експоненційного згладжування та алгоритм для відновлення пропущених значень ряду з використанням цього методу. Створений алгоритм показав перевагу перед методом експоненційного згладжування за значеннями статистичних характеристик під час відновлення місячних значень сонячних радіофлюксів при довжині хвилі 10,7 см за даними Національного управління океанічних і атмосферних досліджень Сполучених Штатів Америки. Розроблено критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів. Метод ковзного двобічного експоненційного згладжування показав перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей за значенням критерію близькості та показав перевагу перед традиційним експоненційним підходом при прогнозуванні за значеннями прогнозних характеристик для даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см.

Розроблено метод адаптивного ковзного двобічного експоненційного згладжування, з його використанням розроблено алгоритм для відновлення пропущених значень часових рядів. За значенням критерію близькості створений метод показав перевагу перед методом 13-місячного ковзного середнього під час відновлення істинних закономірностей чисел Вольфа та місячних значень сонячних радіофлюксів при довжині хвилі 10,7 см. За значеннями статистичних характеристик алгоритм для відновлення пропущених значень часових рядів з використанням методу адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед методом експоненційного згладжування при відновленні пропущених місячних

значень сонячних радіофлюксів при довжині хвилі 10,7 см. За значеннями прогнозних характеристик метод адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед традиційним експоненційним підходом при прогнозуванні місячних значень сонячних радіофлюксів при довжині хвилі 10,7 см.

Визначено принципи системної методології, які застосовані під час розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів. Розроблено архітектуру і структурну схему частини аналізу та прогнозування часових рядів системи підтримки прийняття рішень, створено різні рівні доступу до системи підтримки прийняття рішень в залежності від відділів для співробітників цієї організації. Виконано її програмну реалізацію мовою програмування Java, та створено опис для її користувачів. У створеній системі підтримки прийняття рішень реалізовано можливості виконання аналізу даних, традиційні методи та розроблені нові методи відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Наведено приклади її застосування при роботі з фінансово-економічними та сонячними даними.

У результаті виконання дослідження розроблено та програмно реалізовано цілісну системну методологію для відновлення пропущених значень та істинних закономірностей, прогнозування часових рядів.

Результати дисертації у складі моделей та методів були впроваджені у Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку». Впровадження запропонованої методології дало можливість отримати високоякісні результати відновлення пропущених значень та оцінок прогнозів для взаємозалежних показників сталого розвитку Хорватії.

Результати дисертації у складі моделей, методів та програмних засобів були впроваджені у навчальний процес Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». З використанням запропонованих методів і математичних моделей розроблено лабораторні

роботи для таких дисциплін з навчального плану Інституту прикладного системного аналізу: «Аналіз часових рядів», «Прогнозування у фінансово-економічній сфері» та «Аналіз нелінійних часових рядів». Розроблене програмне та алгоритмічне забезпечення використано для створення сучасної інформаційної технології аналізу фінансових даних, яка взята за основу для лабораторних та курсових робіт з навчальної дисципліни «Системи і методи підтримки прийняття рішень».

Ключові слова: адаптивний фільтр Калмана, метод двобічного експоненційного згладжування, взаємозалежні часові ряди, інтегральний критерій адекватності моделі, критерій близькості, метод ковзного двобічного експоненційного згладжування, метод адаптивного ковзного двобічного експоненційного згладжування, система підтримки прийняття рішень, системний аналіз.

За матеріалами дисертаційного дослідження опубліковано 11 наукових праць: серед них 6 наукових статей у провідних наукових фахових виданнях [1-6] (у тому числі 1 – в іноземному виданні [6]; 1 – в українському виданні, що входить до міжнародної наукометричної бази даних Web of Science [4]; 1 – в українському виданні, що у повному обсязі перевидається англійською мовою під назвою «Journal of Automation and Information Sciences» та входить до міжнародної наукометричної бази даних Scopus [5]), 5 праць у матеріалах доповідей міжнародних конференцій [7-11].

Список публікацій здобувача:

1. Братусь О.В., Подладчиков В.М. Побудова багатовимірної моделі на основі фільтра Калмана й аналіз алгоритмів оцінювання її параметрів. *Наукові вісті НТУУ «КПІ»*. 2013. № 5 (91). С. 28–34.
2. Братусь О.В., Подладчиков В.М. Ідентифікація змінних параметрів моделі для побудови алгоритму прогнозування. *Системні дослідження та інформаційні технології*. 2015. № 3. С. 131–141.

3. Братусь О.В., Подладчиков В.М., Бідюк П.І. Метод двобічного експоненційного згладжування для відновлення динамічних процесів. *Наукові вісті НТУУ «КПІ»*. 2016. № 6 (110). С. 15–21.
4. Братусь О.В. Система підтримки прийняття рішень з адаптивними блоками відновлення та прогнозування сонячних радіофлюксів. *Радіoeлектроніка, інформатика, управління*. 2017. № 3. С. 36–43.
5. Братусь О.В., Бідюк П.І., Болдак А.О. Розробка методів відновлення пропущених значень і прогнозування для взаємозалежних часових рядів. *Міжнародний науково-технічний журнал «Проблеми керування та інформатики»*. 2017. № 5. С. 13–21.
6. Bratus O. Development of adaptive moving two-sided exponential smoothing method for restoring and forecasting of time series. *EUREKA: Physics and Engineering*. 2017. No. 5. P. 13–21.
7. Братусь О.В., Подладчиков В.М. Розробка методів прогнозування на основі фільтра Калмана. *Системний аналіз та інформаційні технології*: матеріали 15-ї Міжнародної науково-технічної конференції SAIT 2013, м. Київ, ННК «ІПСА» НТУУ «КПІ», 27-31 травня 2013 р. Київ: ННК «ІПСА» НТУУ «КПІ», 2013. С. 251.
8. Братусь О.В., Подладчиков В.М. Пошук закономірностей динамічних процесів на основі двобічного експоненційного згладжування. *Контроль і управління в складних системах (КУСС-2014)*: тези доповідей XII міжнародної конференції, м. Вінниця, 14-16 жовтня 2014 р. Вінниця: ВНТУ, 2014. С. 28.
9. Братусь О.В. Система підтримки прийняття рішень для аналізу та прогнозування цін на метали. *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту*: матеріали міжнародної наукової конференції, с. Залізний Порт, 24-28 травня 2016 р. Херсон: ПП Вишемирський В.С., 2016. С. 185–186.
10. Братусь О.В. Ідентифікація змінної дисперсії шуму вимірів моделі у задачах прогнозування. *Математика. Інформаційні технології. Освіта*: тези доповідей

V міжнародної конференції, м. Луцьк - с. Світязь, 5-7 червня 2016 р. Луцьк: ПП Іванюк В.П., 2016. С. 58–59.

11. Братусь О.В., Подладчиков В.М., Подладчикова Т.В. Метод ковзного двобічного експоненційного згладжування для відновлення закономірностей динамічних процесів. *Математичне та імітаційне моделювання систем. МОДС 2016*: тези доповідей одинадцятої міжнародної науково-практичної конференції, с. Жукин, 27 червня - 1 липня 2016 р. Чернігів: ЧНТУ, 2016. С. 28–31.

ABSTRACT

Bratus O.V. Methods of forecasting of non-stationary time series based on two-sided exponential smoothing and optimal filtering. – Qualifying scientific work on the rights of manuscript.

Dissertation for scientific degree of candidate of technical sciences on the speciality 01.05.04 – system analysis and optimal decisions theory (technical sciences). – National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ministry of education and science of Ukraine. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ministry of education and science of Ukraine, Kyiv, 2019.

The dissertation work is dedicated to development and application of applied scientific methodology of system analysis for solving of actual tasks of restoring of missing values of time series, restoring of true regularities of researched processes, forecasting of time series and performing of automated processing of time series using created decision making support system.

Restoring of missing values and true regularities are important tasks nowadays, because it takes ability to perform correctly previous processing and further forecasting of time series. It is important to get more accurate forecasting estimates of time series values for those, who are working on forecasting of time series in their everyday work. Development of new decision making support systems for analysis

and forecasting of time series based on modern principles of system analysis is important, because such improved systems will give ability to perform fast and more accurate processing of data.

Methods of estimation of mathematical expectation of acceleration of values change of data samples using full and rarified samples of residuals of suboptimal filter with memory two were developed in this work, they were compared with existing methods. The best methods were selected between developed methods, and they were used for development of adaptive Kalman filters. Developed adaptive Kalman filters showed better results in comparison to exponential smoothing, autoregressive, autoregressive moving average methods and group method of data handling in forecasting of daily average prices of zinc by the London metal exchange data.

Method of estimation of mathematical expectation of acceleration of changes of time series values, which is changed based on the unknown law, using the exponential smoothing procedure to the constructed series of one-dependent pseudo-measurements of this parameter was developed, and forecasting algorithms based on Kalman filter were created using developed method. Created forecasting algorithms based on Kalman filter showed superiority in comparison to autoregressive, autoregressive moving average methods and group method of data handling by forecasting characteristics in forecasting of daily average prices of lead by the London metal exchange data.

Two-sided exponential smoothing method and algorithm for restoring of missing values of time series using this method were developed. The integral criterion of model adequacy was created. Application of the created method for restoring of true regularities and forecasting of time series was described. The two-sided exponential smoothing method showed superiority in comparison to arithmetic mean, autoregressive approach and exponential smoothing method by statistical characteristics in restoring of missing values of daily average prices of zinc by the London metal exchange data.

Method for restoring missing values for mutually dependent time series using two-sided exponential smoothing was developed. This method showed superiority in comparison to arithmetic mean and exponential smoothing methods by statistical characteristics in restoring mutually dependent indices of sustainable development by the data of World data center. Method for forecasting of mutually dependent time series using two-sided exponential smoothing method was developed. This method showed superiority in comparison to exponential smoothing method by the forecasting characteristics in forecasting of sustainable development indices.

Moving two-sided exponential smoothing method and algorithm for restoring of missing values of time series using this method were developed. Created algorithm showed superiority in comparison to exponential smoothing method by statistical characteristics in restoring of monthly values of solar radio fluxes at a wavelength of 10,7 cm by the data of National oceanic and atmospheric administration of the United States of America. The proximity criterion for estimation of models in restoring of the true regularities of time series was developed. Moving two-sided exponential smoothing method showed superiority in comparison to 13-month moving average method in restoring of the true regularities by the value of the proximity criterion and showed superiority in comparison to traditional exponential approach by forecasting characteristics in forecasting of monthly values of solar radio fluxes at a wavelength of 10,7 cm.

Adaptive moving two-sided exponential smoothing method was developed, algorithm for restoring of missing values of time series using this method was developed. The created method showed superiority by proximity criterion in comparison to 13-month moving average method in restoring of the true regularities of Wolf numbers and monthly values of solar radio fluxes at a wavelength of 10,7 cm. Algorithm for restoring of missing values of time series using adaptive moving two-sided exponential smoothing method showed superiority by statistical characteristics values in comparison to exponential smoothing method in restoring of missing values of monthly values of solar radio fluxes at a wavelength of 10,7 cm. Adaptive moving two-sided exponential smoothing method showed superiority by

forecasting characteristics in comparison to traditional exponential approach in forecasting of monthly values of solar radio fluxes at a wavelength of 10,7 cm.

Principles of system methodology were determined, which were used during development of decision making support system for organization, which is working on analysis and forecasting of time series. Architecture and structural scheme of part of analysis and forecasting of decision making support system were developed, different levels of access to the decision making support system were created according to departments for employees of this organization. Software implementation of it was performed using Java programming language, and description for its users was created. Opportunities of data analysis performing, traditional methods and created new methods of restoring of missing values, restoring of true regularities and forecasting of time series were implemented in the created decision making support system. Examples of its using for work with financial-economic and solar data were provided.

As a result of performing the research, holistic system methodology was developed, and software was implemented for this methodology for restoring of missing values and true regularities, forecasting of time series.

Results of dissertation consisting of models and methods were introduced at the Educational-scientific complex “World data center for geoinformatics and sustainable development”. Introducing of proposed methodology was given ability to reach high quality results of restoring of missing values and forecasting estimations for mutually dependent indices of sustainable development of Croatia.

Results of dissertation consisting of models, methods and software were introduced into the study process of Institute for applied system analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”. Laboratory works were developed using created methods and mathematical models for the following disciplines of study plan of Institute for applied system analysis: “Time series analysis”, “Forecasting in the financial-economic sphere” and “Analysis of non-linear time series”. Created software and algorithms were used for development of modern information technology of financial data analysis, which was

selected as basis for laboratory and course works for study discipline “Systems and methods of decision making”.

Key words: adaptive Kalman filter, two-sided exponential smoothing method, mutually dependent time series, integral criterion of model adequacy, proximity criterion, moving two-sided exponential smoothing method, adaptive moving two-sided exponential smoothing method, decision making support system, system analysis.

11 scientific papers were published based on the materials of dissertation research: 6 scientific articles of them in the leading scientific editions [1-6] (including 1 – in the foreign edition [6]; 1 – in the Ukrainian edition, which is included into the international scientometric database Web of Science [4]; 1 – in the Ukrainian edition, which full content is republished in English named «Journal of Automation and Information Sciences» and included into the international scientometric database Scopus [5]), 5 papers in the materials of reports of international conferences [7-11].

Publications:

1. Bratus O.V., Podladchikov V.M. Construction of multidimensional models based on Kalman filter and analysis of algorithms of estimation of its parameters. *Research Bulletin of the NTUU “KPI”*. 2013. Vol. 91, No. 5. P. 28–34.
2. Bratus O.V., Podladchikov V.M. Identification of variable parameters of a model for the construction of a forecasting algorithm. *System research and information technologies*. 2015. No. 3. P. 131–141.
3. Bratus O.V., Podladchikov V.M., Biduyk P.I. Two-sided exponential smoothing method for restoring of dynamic processes. *Research Bulletin of the NTUU “KPI”*. 2016. Vol. 110, No. 6. P. 15–21.
4. Bratus O.V. Decision making support system with adaptive units of restoring and forecasting of solar radio fluxes. *Radio Electronics, Computer Science, Control*. 2017. No. 3. P. 36–43.
5. Bratus O.V., Bidyuk P.I., Boldak A.O. Development of methods for restoring missing values and forecasting of mutually dependent time series. *International*

- Scientific Technical Journal "Problems of Control and Informatics"*. 2017. No. 5. P. 13–21.
6. Bratus O. Development of adaptive moving two-sided exponential smoothing method for restoring and forecasting of time series. *EUREKA: Physics and Engineering*. 2017. No. 5. P. 13–21.
 7. Bratus O.V., Podladchikov V.M. Development of forecasting methods based on Kalman filter. *System analysis and information technologies: proceedings of the 15-th International scientific-technical conference SAIT 2013, Kyiv, ESC "IASA" NTUU "KPI", May 27-31, 2013. Kyiv: ESC "IASA" NTUU "KPI", 2013. P. 251.*
 8. Bratus O.V., Podladchikov V.M. Search of regularities of dynamic processes based on two-sided exponential smoothing. *Measurement and control in complex systems (MCCS-2014): abstracts of the XII international conference, Vinnytsia, October 14-16, 2014. Vinnytsia: VNTU, 2014. P. 28.*
 9. Bratus O.V. Decision making support system for analysis and forecasting of metals prices. *Intellectual systems for decision making and problems of computational intelligence: proceedings of international scientific conference, Zaliznyj Port, May 24-28, 2016. Kherson: PP Vyshemirsky V.S., 2016. P. 185–186.*
 10. Bratus O.V. Identification of variable dispersion of measurements noise in forecasting tasks. *Math. Information technologies. Education: abstracts of the V international conference, Lutsk-Svityaz, June 5-7, 2016. Lutsk: PP Ivanyuk V.P., 2016. P. 58–59.*
 11. Bratus O.V., Podladchikov V.M., Podladchikova T.V. Moving two-sided exponential smoothing method for restoring of regularities of dynamic processes. *Mathematical modelling and simulation of systems. MODS 2016: abstracts of the eleventh international scientific-practical conference, Zhukin, June 27 - July 1, 2016. Chernihiv: CNTU, 2016. P. 28–31.*

ЗМІСТ

АНОТАЦІЯ	2
ПЕРЕЛІК СКОРОЧЕНЬ	18
ВСТУП	20
1. ОГЛЯД СУЧАСНИХ МЕТОДІВ АНАЛІЗУ, ВІДНОВЛЕННЯ ІСТИННИХ ЗАКОНОМІРНОСТЕЙ, ПРОПУЩЕНИХ ЗНАЧЕНЬ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ	29
1.1. Застосування системного підходу до аналізу та прогнозування часових рядів.....	30
1.2. Аналіз особливостей динаміки процесів, що досліджуються.....	32
1.2.1. Наявність нестационарності: існування трендів та гетероскедастичності.....	32
1.2.2. Наявність сезонних ефектів.....	35
1.2.3. Методи відновлення пропущених значень часових рядів.....	36
1.3. Огляд методів ідентифікації моделей динаміки нестационарних процесів.....	38
1.3.1. Метод 13-місячного ковзного середнього.....	38
1.3.2. Модель сонячного циклу з використанням визначення параметрів сонячного циклу.....	39
1.3.3. Використання різницьових рівнянь для опису статистичних даних.....	40
1.4. Методи і моделі для прогнозування часових рядів.....	41
1.4.1. Регресійні моделі.....	42
1.4.2. Методи експоненційного згладжування.....	43
1.4.3. Фільтр Калмана.....	45
1.4.4. Метод групового урахування аргументів.....	49
1.5. Застосування систем підтримки прийняття рішень для аналізу та прогнозування часових рядів.....	52
1.6. Висновки до розділу 1 та постановка задач дослідження.....	55
2. РОЗРОБКА МЕТОДІВ ОЦІНЮВАННЯ НЕВІДОМИХ ПАРАМЕТРІВ МАТЕМАТИЧНОЇ МОДЕЛІ ТА ПОБУДОВА АДАПТИВНИХ ФІЛЬТРІВ КАЛМАНА З ВИКОРИСТАННЯМ ОТРИМАНИХ ОЦІНОК	59

2.1. Застосування системного підходу до аналізу часових рядів та розробки методів їх прогнозування.....	60
2.2. Аналіз методів оцінювання параметрів математичної моделі та побудова адаптивних фільтрів Калмана з використанням обчислених оцінок.....	62
2.2.1. Математична постановка задачі.....	63
2.2.2. Побудова та аналіз методів оцінювання математичного сподівання прискорення зміни значень вибірки даних.....	65
2.2.3. Розробка адаптивних фільтрів Калмана та їх застосування для прогнозування середньодобових цін на цинк.....	70
2.3. Побудова алгоритмів прогнозування на основі фільтра Калмана з використанням методів ідентифікації змінних параметрів моделі.....	75
2.3.1. Математична постановка задачі.....	77
2.3.2. Метод ідентифікації математичного сподівання прискорення.....	77
2.3.3. Імітаційне моделювання розробленого методу ідентифікації.....	79
2.3.4. Алгоритми прогнозування на основі фільтра Калмана.....	80
2.3.5. Прогнозування середньодобових цін на свинець.....	86
2.4. Висновки до розділу 2.....	88
3. РОЗРОБКА МЕТОДІВ ВІДНОВЛЕННЯ ПРОПУЩЕНИХ ЗНАЧЕНЬ, ІСТИННИХ ЗАКОНОМІРНОСТЕЙ ТА ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ НА ОСНОВІ ДВОБІЧНОГО ЕКСПОНЕНЦІЙНОГО ЗГЛАДЖУВАННЯ.....	90
3.1. Розробка методу двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.....	91
3.1.1. Опис методу двобічного експоненційного згладжування.....	92
3.1.2. Імітаційне моделювання для дослідження ефективності методу двобічного експоненційного згладжування.....	95
3.1.3. Застосування методу двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.....	98

3.1.4. Відновлення пропущених значень цін на цинк.....	99
3.2. Розробка методів відновлення пропущених значень та прогнозування для взаємозалежних часових рядів.....	103
3.2.1. Метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.....	104
3.2.2. Приклади застосування методу відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.....	107
3.2.3. Метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.....	112
3.2.4. Приклад застосування методу прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування...	113
3.3. Розробка методу ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.....	114
3.3.1. Метод ковзного двобічного експоненційного згладжування.....	116
3.3.2. Приклади застосування методу ковзного двобічного експоненційного згладжування для відновлення пропущених значень та істинних закономірностей часових рядів.....	119
3.3.3. Приклад застосування методу ковзного двобічного експоненційного згладжування для прогнозування часових рядів.....	124
3.4. Розробка методу адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.....	125
3.4.1. Метод адаптивного ковзного двобічного експоненційного згладжування.....	127
3.4.2. Застосування методу адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.....	129

3.4.3. Приклади відновлення пропущених значень та істинних закономірностей часових рядів з використанням методу адаптивного ковзного двобічного експоненційного згладжування.....	131
3.4.4. Приклад прогнозування значень часових рядів з використанням методу адаптивного ковзного двобічного експоненційного згладжування.....	137
3.5. Висновки до розділу 3.....	138
4. РОЗРОБКА СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОРГАНІЗАЦІЇ, ЩО ЗАЙМАЄТЬСЯ АНАЛІЗОМ ТА ПРОГНОЗУВАННЯМ ЧАСОВИХ РЯДІВ.....	143
4.1. Застосування системного підходу до розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів.....	144
4.1.1. Бізнес-правила.....	147
4.1.2. Матриці подій для співробітників відділів організації.....	148
4.2. Проектування бази даних.....	149
4.3. Опис програмної реалізації системи підтримки прийняття рішень.....	151
4.4. Висновки до розділу 4.....	174
ВИСНОВКИ.....	177
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	181
Додаток А. Акт впровадження результатів дисертаційної роботи у Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку».....	190
Додаток Б. Довідка про впровадження результатів дисертаційної роботи у навчальний процес Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».....	191
Додаток В. Список публікацій здобувача та апробація результатів дисертації.....	192

ПЕРЕЛІК СКОРОЧЕНЬ

АКДЕЗ – адаптивне ковзне двобічне експоненційне згладжування

АКФ – автокореляційна функція

АПОФК – алгоритм прогнозування на основі фільтра Калмана

АР – авторегресія

АРІКС – авторегресія інтегрована з ковзним середнім

АРКС – авторегресія з ковзним середнім

АРП – авторегресійний підхід

АРУГ – авторегресія умовно гетероскедастична

АТО – антитерористична операція

АФК – адаптивний фільтр Калмана

БД – база даних

д/б НДР – державна/бюджетна науково-дослідна робота

ДЕЗ – двобічне експоненційне згладжування

ДР – державна реєстрація

ЕЗ – експоненційне згладжування

ЕП – експоненційний підхід

КДЕЗ – ковзне двобічне експоненційне згладжування

КС – ковзне середнє

МВЗВЧР – метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування

МГУА – метод групового урахування аргументів

МНК – метод найменших квадратів

МПВЧР – метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування

СА – середнє арифметичне

СППР – система підтримки прийняття рішень

США – Сполучені Штати Америки

ТКС – 13-місячне ковзне середнє

УАРУГ – узагальнена авторегресія умовно гетероскедастична

УАРУГ-М – узагальнена авторегресія умовно гетероскедастична модифікована

ЧАКФ – часткова автокореляційна функція

DW – статистика Дарбіна-Уотсона (англ. Durbin-Watson)

ER-діаграма – діаграма «сутність-зв'язок» (англ. entity-relationship diagram)

MAPE – середня відсоткова абсолютна похибка (англ. Mean absolute percentage error)

N – подія нормального типу у матриці подій

RMSE – середньоквадратична похибка (англ. Root mean square error)

SFU – одиниця виміру сонячних радіофлюксів (англ. Solar Flux Unit)

SSE – сума квадратів похибок (англ. Sum squared error)

ВСТУП

Обґрунтування вибору теми дослідження

Дисертаційна робота присвячена розробці та застосуванню прикладної наукової методології системного аналізу для розв'язання задач відновлення пропущених значень часових рядів, відновлення істинних закономірностей розвитку досліджуваних процесів, прогнозування часових рядів та виконання автоматизованої обробки часових рядів за допомогою створеної системи підтримки прийняття рішень. Відновлення пропущених значень часових рядів є важливою проблемою на сьогоднішній день, тому що це дає змогу правильно виконати попередню обробку часових рядів та їх подальше прогнозування. Відновлення істинних закономірностей досліджуваних процесів є актуальним для часових рядів, що містять значні шумові складові та потребують попередньої обробки перед побудовою математичної моделі та прогнозуванням. Отримання точніших оцінок значень часових рядів є важливим для тих, хто займається прогнозуванням часових рядів у своїй повсякденній роботі, наприклад, займається купівлею та продажем металів на біржі металів або інших активів. Розробка нових систем підтримки прийняття рішень для аналізу та прогнозування часових рядів на основі сучасних принципів системного аналізу є важливою задачею, оскільки такі вдосконалені системи дадуть змогу виконувати швидко та більш точну обробку даних.

Важливі теоретичні дослідження у галузі системного аналізу виконані багатьма видатними вченими: академіком В.І. Вернадським, О.О. Богдановим, Л. фон Берталанфі, Т. Котарбінським, Н. Вінером. Ці дослідження були у подальшому розвинуті у сучасних роботах академіка М.З. Згуровського, член-кореспондента Н.Д. Панкратової, В.Д. Романенка, В.Я. Данилова, В.О. Капустяна, О.А. Павлова, Т. Сааті, О.В. Антонова, Н.З. Шора, Р.Л. Акоффа та інших вчених.

Відомими сучасними вченими у галузі аналізу та прогнозування часових рядів є Ю.П. Зайченко, В.Д. Романенко, К.Т. Леондес, Дж. Бендат, А. Пірсол.

Докладні дослідження стосовно фільтра Калмана виконано у роботах Р. Мехра, В.М. Подладчікова. Важливою задачею є також ідентифікація параметрів систем для подальшого їх використання при прогнозуванні. Дослідження стосовно методів ідентифікації параметрів систем виконані у роботах К.Й. Острема та П. Ейкоффа. Е.С. Гарднер молодший виконав дослідження стосовно експоненційного згладжування та створив класифікацію методів експоненційного згладжування. Дослідженням відновлення істинних закономірностей процесів та прогнозування сонячних даних займалися багато науковців. У роботах Д.Г. Хетеуея, Р.М. Вілсона та Е.Дж. Рейхмана запропоновано рівняння для опису моделі 11-річного сонячного циклу, яке відповідно до їх досліджень може бути використаним для прогнозування. У подальших роботах цих авторів запропоновано метод розрахунку 13-місячного ковзного середнього для обробки даних про сонячну активність, а також виконано докладний огляд багатьох існуючих методів прогнозування сонячних даних.

Існує значна кількість нерозв'язаних задач, що стосуються аналізу та прогнозування часових рядів та є актуальними на сьогоднішній день: підвищення адекватності математичних моделей часових рядів; отримання точніших оцінок значень часових рядів під час відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Дисертаційна робота присвячена виконанню досліджень стосовно розв'язання актуальних наведених задач, а також розробці оригінального програмного продукту у формі системи підтримки прийняття рішень для забезпечення автоматизованого розв'язання цих задач з використанням традиційних та розроблених у роботі методів.

Зв'язок роботи з науковими програмами, планами, темами

Дисертаційна робота виконана на кафедрі математичних методів системного аналізу Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» у відповідності до планів науково-дослідних робіт:

1. д/б НДР «Розробка інформаційної технології моделювання та оцінювання фінансово-економічних ризиків із врахуванням невизначеностей різної природи (на основі байєсівських моделей)» (№ ДР 0113U000650);
2. д/б НДР «Розробка методології системного аналізу, моделювання та оцінювання фінансових ризиків» (№ ДР 0115U000356);
3. д/б НДР «Проектування сучасних систем сервісів на прикладі мобільної медичної системи для мешканців прифронтових селищ в зоні АТО» (№ ДР 0117U002435).

Мета і завдання дослідження

Мета дослідження – розробка методів оцінювання параметрів математичних моделей, що змінюються за невідомим законом, та їх застосування для підвищення якості оцінок прогнозів розвитку часових рядів шляхом удосконалення системної методології прогнозування; розробка методів відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробка оригінального програмного продукту у формі системи підтримки прийняття рішень. Для досягнення мети потрібно вирішити такі завдання:

- для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створити методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два; обрати кращі методи серед створених та застосувати їх для розробки адаптивних фільтрів Калмана;

- для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створити метод оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду незалежних псевдовимірів цього параметру; застосувати створений метод для розробки алгоритмів прогнозування на основі фільтра Калмана;

- розробити метод двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробити інтегральний критерій адекватності моделі для вибору оптимального параметра згладжування з метою застосування у методах, які пов’язані з експоненційним згладжуванням;

- розробити методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування;

- розробити метод ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробити критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів;

- розробити метод адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів;

- створити архітектуру та структурну схему системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів, а також виконати її програмну реалізацію мовою програмування Java.

Об’єкт дослідження: процеси, що описуються фінансово-економічними часовими рядами, показниками сталого розвитку, сонячними даними (числами Вольфа, сонячними радіофлюксами при довжині хвилі 10,7 см).

Предмет дослідження: моделі і методи відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.

Методи дослідження

У роботі використано такі методи дослідження: методи регресійного аналізу, адаптивний регресійний підхід, відновлення пропущених значень часових рядів за допомогою середнього арифметичного, метод групового урахування аргументів, методи оцінювання математичного сподівання прискорення зміни значень вибірки даних, адаптивні фільтри Калмана,

алгоритми прогнозування на основі фільтра Калмана, метод експоненційного згладжування, метод подвійного експоненційного згладжування, експоненційний підхід до опису сонячного циклу, метод 13-місячного ковзного середнього, метод двобічного експоненційного згладжування, метод ковзного двобічного експоненційного згладжування, метод адаптивного ковзного двобічного експоненційного згладжування, методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.

Наукова новизна одержаних результатів

Виконані у дисертаційній роботі дослідження дозволили отримати такі теоретичні та практичні результати:

Уперше:

- для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, запропоновані методи дають можливість оцінити цей невідомий параметр;

- розроблено метод двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів, що дає змогу отримувати точніші оцінки часового ряду за рахунок врахування інформації про часовий ряд, отриманої у зворотному часі; розроблено інтегральний критерій адекватності моделі для вибору оптимального значення параметра згладжування, який відрізняється одночасним врахуванням двох статистик і забезпечує вибір адекватної моделі;

- розроблено методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування, ці методи дають можливість враховувати динаміку взаємозалежних часових рядів для отримання точніших оцінок значень часових рядів;

– розроблено методи ковзного двобічного експоненційного згладжування та адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів, ці методи призначені для роботи з сонячними даними і дають можливість враховувати специфіку таких даних.

Удосконалено:

– адаптивні фільтри Калмана та алгоритми прогнозування на основі фільтра Калмана з використанням методів оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, які відрізняються врахуванням оцінок невідомого математичного сподівання прискорення зміни значень вибірки даних і дають можливість отримати точніші оцінки прогнозів часових рядів.

Отримав подальший розвиток:

– критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів.

Практичне значення одержаних результатів

1. У результаті виконання дослідження розроблено та програмно реалізовано цілісну системну методологію для відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Розроблена система підтримки прийняття рішень застосована для аналізу, відновлення пропущених значень, відновлення істинних закономірностей та прогнозування фінансово-економічних та сонячних даних.

2. Результати дисертації у складі моделей та методів впроваджені у Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» (акт впровадження № 0301/18 від 13.03.2018, наведено у додатку А). Впровадження запропонованої методології дало можливість отримати високоякісні результати відновлення пропущених значень та оцінок прогнозів для взаємозалежних показників сталого розвитку Хорватії (багатовимірні часові ряди).

3. Результати дисертації у складі моделей, методів та програмних засобів були впроваджені у навчальний процес Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (довідка про впровадження № 48/4040 від 22.03.2018, наведено у додатку Б). З використанням запропонованих методів і математичних моделей розроблено лабораторні роботи для таких дисциплін з навчального плану Інституту прикладного системного аналізу: «Аналіз часових рядів», «Прогнозування у фінансово-економічній сфері» та «Аналіз нелінійних часових рядів». Розроблене програмне та алгоритмічне забезпечення використано для створення сучасної інформаційної технології аналізу фінансових даних, яка взята за основу для лабораторних та курсових робіт з навчальної дисципліни «Системи і методи підтримки прийняття рішень».

Особистий внесок здобувача

Всі основні наукові положення та результати, що складають основний зміст роботи та становлять наукову новизну, отримані автором самостійно. Список публікацій та апробацію результатів дисертації наведено у додатку В.

У роботі [1], опублікованій у співавторстві, здобувачеві належать створені методи № 4–6 для оцінювання математичного сподівання прискорення зміни значень вибірки даних, використання розробленого методу № 5 для побудови адаптивного фільтра Калмана та подальшого прогнозування цін на цинк.

У роботі [2], опублікованій у співавторстві, здобувачеві належать метод оцінювання математичного сподівання прискорення зміни значень вибірки даних та створений алгоритм прогнозування на основі фільтра Калмана з використанням цього методу.

У роботі [3], опублікованій у співавторстві, здобувачеві належать розроблені метод двобічного експоненційного згладжування та інтегральний критерій адекватності моделі.

У роботі [5], опублікованій у співавторстві, здобувачеві належать метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування та метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.

У доповіді [7], опублікованій у співавторстві, здобувачеві належать методи оцінювання математичного сподівання прискорення зміни значень вибірки даних та адаптивний фільтр Калмана.

У доповіді [8], опублікованій у співавторстві, здобувачеві належить метод двобічного експоненційного згладжування.

У доповіді [11], опублікованій у співавторстві, здобувачеві належать метод ковзного двобічного експоненційного згладжування та критерій близькості.

Апробація результатів дисертації

Основні положення дисертаційної роботи були представлені на наукових конференціях та семінарах:

- 15^й міжнародній науково-технічній конференції SAIT 2013 (Україна, м. Київ, 27-31 травня 2013 р.);
- XII міжнародній конференції «Контроль і управління в складних системах (КУСС-2014)» (Україна, м. Вінниця, 14-16 жовтня 2014 р.);
- міжнародній науковій конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (Україна, с. Залізний Порт, 24-28 травня 2016 р.);
- V^й міжнародній конференції «Математика. Інформаційні технології. Освіта» (Україна, м. Луцьк - с. Світязь, 5-7 червня 2016 р.);
- Одинадцятій міжнародній науково-практичній конференції «Математичне та імітаційне моделювання систем. МОДС 2016» (Україна, с. Жукин, 27 червня - 1 липня 2016 р.);
- науковому семінарі «Системні дослідження та інформаційні технології» в Інституті прикладного системного аналізу Національного технічного

університету України «Київський політехнічний інститут імені Ігоря Сікорського» (Україна, м. Київ, 21 березня 2018 р.).

Публікації

За матеріалами дисертаційного дослідження опубліковано 11 наукових праць: серед них 6 наукових статей у провідних наукових фахових виданнях [1-6] (у тому числі 1 – в іноземному виданні [6]; 1 – в українському виданні, що входить до міжнародної наукометричної бази даних Web of Science [4]; 1 – в українському виданні, що у повному обсязі перевидається англійською мовою під назвою «Journal of Automation and Information Sciences» та входить до міжнародної наукометричної бази даних Scopus [5]), 5 праць у матеріалах доповідей міжнародних конференцій [7-11].

Структура та обсяг дисертації

Дисертаційна робота складається з анотації, змісту, переліку скорочень, вступу, чотирьох основних розділів, висновків, списку використаних джерел, трьох додатків. Робота викладена на 194 сторінках і містить 161 сторінку основної частини, 34 рисунки, 21 таблицю, список використаних джерел із 89 найменувань на 9 сторінках, 3 додатки на 5 сторінках.

РОЗДІЛ 1

ОГЛЯД СУЧАСНИХ МЕТОДІВ АНАЛІЗУ, ВІДНОВЛЕННЯ ІСТИННИХ ЗАКОНОМІРНОСТЕЙ, ПРОПУЩЕНИХ ЗНАЧЕНЬ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Аналіз особливостей динаміки часових рядів є важливим на сьогоднішній день, тому що попередній аналіз дає змогу визначити характерні особливості часових рядів, що досліджуються. Важливою є розробка та удосконалення системної методології аналізу даних у формі часових рядів, яка забезпечує коректний підхід до розв'язання задач моделювання, прогнозування та автоматизованої обробки часових рядів.

Аналіз даних на наявність нестационарності дає можливість у випадку її виявлення прийняти подальше рішення стосовно врахування гетероскедастичності, тренду або виконання подальшої обробки ряду з видаленням тренду. Врахування сезонних ефектів для часових рядів допомагає більш точно виконувати обробку та прогнозування часового ряду з використанням інформації про сезонний вплив певних факторів. Відновлення пропущених значень часових рядів дає змогу працювати з повним часовим рядом без видалення проміжків, на яких відсутня інформація про ряд. Це забезпечує виконання дослідження без втрати інформації.

Важливим для подальшої обробки та прогнозування часового ряду є знаходження істинних закономірностей ряду, особливо для специфічних даних, таких, як дані про сонячну активність. Тому розробка та вдосконалення методів відновлення істинних закономірностей часових рядів є актуальними на сьогоднішній день.

Розробка та вдосконалення методів прогнозування часових рядів є актуальними, тому що це дасть змогу отримувати точніші прогнозні оцінки значень часових рядів.

Важливою задачею на сьогоднішній день є створення систем підтримки прийняття рішень для аналізу та прогнозування часових рядів, оскільки

розробка таких систем дозволяє виконувати швидку та високоякісну автоматизовану обробку часових рядів.

Таким чином, удосконалення системної методології, створення нових методів та вдосконалення існуючих методів для аналізу та прогнозування часових рядів, розробка систем підтримки прийняття рішень для автоматизованої обробки часових рядів є актуальними завданнями на сьогоднішній день.

1.1. Застосування системного підходу до аналізу та прогнозування часових рядів

Аналіз даних у формі часових рядів є системною задачею, оскільки він має всі ознаки такої задачі. У роботі [12] показано, що системною задачею є задача аналізу певної сукупності властивостей об'єкта дослідження з єдиної позиції цілісного підходу для досягнення заданих цілей за наявних умов. Процедура аналізу та прогнозування значень часового ряду можна вважати процедурою розв'язання формалізованої задачі системного аналізу. Процедуру розв'язання формалізованих задач здійснюють за два етапи [12]:

- на першому етапі виконується формалізація задачі і підготовка відповідних матеріалів до обчислень;
- на другому етапі обчислення проводяться на комп'ютері або в комп'ютерній мережі.

Принцип декомпозиції вихідної задачі на послідовність більш простих із подальшим агрегуванням результатів розв'язання окремих задач у підсумковий розв'язок вихідної загальної задачі системного аналізу застосовується для розв'язання великої кількості задач системного аналізу [13]. Процес декомпозиції триває до такого рівня, коли отримана часткова задача має формалізовану частину, яку можна розв'язати за допомогою відомих методів.

Задачам системного аналізу властиві багатокритеріальність, багатofакторність і багатопараметричність [12]. Ці властивості описуються такими характеристиками [12]:

- критерій – це показник або функціонал, який дає змогу кількісно або якісно оцінювати ту чи іншу властивість системи;
- фактор – зовнішній або внутрішній вплив, що діє на певні властивості та поведінку складних систем;
- параметр – кількісний показник, який визначає деяку властивість системи або ступінь впливу фактора на її властивість.

Під час виконання аналізу та прогнозування часових рядів застосовується певний набір критеріїв, що характеризує адекватність моделі у цілому та необхідний для підтвердження застосування моделі, що розглядається, для прогнозування. До факторів, що впливають на часові ряди, можна віднести зовнішні збурення стану та шуми вимірів, сезонні ефекти, екстремальні значення. Параметри систем найчастіше є невизначеними і потребують застосування додаткових процедур ідентифікації.

Під час створення нових методів для аналізу та прогнозування часових рядів необхідно враховувати, що ці методи повинні задовольняти таким фундаментальним властивостям системної методології, як результативність, ефективність та масштабність. Ці властивості докладно описані у роботі [12].

Необхідно звернути увагу на рівень обчислювальної складності, ступінь структурованості і рівень потенційної формалізованості задачі під час виконання системних досліджень відповідно до підходу Дж. Кліра [14]. Ці фактори повинні бути враховані під час створення нових методів для аналізу та прогнозування часових рядів, відповідних систем підтримки прийняття рішень.

Під час розробки нових методів аналізу та прогнозування часових рядів, а також систем підтримки прийняття рішень, в яких ці методи реалізуються, необхідно дотримуватись принципів формування системної методології та її практичної реалізації. Фундаментальними принципами формування системної методології та її практичної реалізації є такі [12]: принцип системної

погодженості; принцип процедурної повноти; принцип функціональної ортогональності; принцип інформаційної взаємозалежності; принцип цілеспрямованої відповідності; принцип функціональної раціональності; принцип багатоцільової загальності; принцип багатофакторної адаптивності; принцип процедурної відкритості; принцип раціональної доповнюваності.

Таким чином, аналіз та прогнозування часових рядів є системними задачами. Під час розробки нових методів для аналізу та прогнозування часових рядів, а також систем підтримки прийняття рішень, в яких ці методи повинні бути реалізованими, необхідно дотримуватись фундаментальних принципів формування системної методології та її практичної реалізації для досягнення найкращих результатів та розробки ефективних методів.

1.2. Аналіз особливостей динаміки процесів, що досліджуються

Аналіз особливостей динаміки процесів, що досліджуються, є важливим на сьогоднішній день, оскільки він дозволяє отримати інформацію, яка може бути використаною при подальшій роботі з часовим рядом, що описує процес. Важливим є аналіз на наявність нестационарності, сезонних ефектів, пропущених значень з можливістю їх відновлення.

1.2.1. Наявність нестационарності: існування трендів та гетероскедастичності

Аналіз часового ряду починається з аналізу на наявність тренду. Наявність тренду можна помітити, проаналізувавши графік часового ряду.

У роботі [15] тренд визначається як поточне середнє значення ряду, яке вказує на довгострокову зміну процесу. Поширеним є також означення тренду як тривалої зміни середнього рівня випадкового процесу [16]. У загальному випадку існують такі задачі аналізу часового ряду з трендом [16]:

- обчислення тренду та усунення випадкових відхилень;

- обчислення та усунення тренду з метою дослідження відхилень.

Для опису тренду використовується поширений підхід, що полягає у використанні детермінованої функції від часу такого вигляду [17]:

$$y(k) = a_0 + d_1 k + d_2 k^2, \quad (1.1)$$

де a_0 – числовий коефіцієнт (зміщення); k – дискретний час; d_i – коефіцієнти рівняння. Тренд, який описується такою функцією, називається детермінованим або глобальним трендом [17].

На сьогоднішній день існує тенденція формування більш загального підходу до опису тренду, зокрема, замість глобальних моделей використовуються локальні [18]. Тренд розглядається при цьому як стохастична функція від часу [18]. Методи опису тренду та підходи до обробки рядів, що містять тренд, наведено у роботі [19]. Для опису локального тренду використовується введення залежності коефіцієнтів моделі від часу [20]:

$$y(k) = a(k) + d_1(k)k, \quad (1.2)$$

де $a(k)$ – локальна константа; $d_1(k)$ – коефіцієнт, що визначає локальний нахил тренду. У роботі [18] показано, що функції вигляду (1.2) є більш робастними, ніж функції вигляду (1.1).

Існує альтернативний підхід до опису локального тренду. Він полягає у використанні рекурсивних рівнянь такого вигляду [18]:

$$y(k) = a_0 + y(k-1), \quad (1.3)$$

або в ускладненому варіанті рівняння такого вигляду [18]:

$$y(k) = a_0(k) + y(k-1) + \varepsilon(k-1), \quad (1.4)$$

де $\varepsilon(k-1)$ – випадкова змінна, яку для простоти часто вважають послідовністю білого шуму з відомою дисперсією.

У роботі [21] виконано дослідження стосовно визначення тренду для часових рядів з використанням внутрішнього розкладання за часовою шкалою.

Також для опису тренду використовується модель випадкового кроку з дрейфом (зміщенням або перетином) [18]. У такому випадку додається константа a_0 до моделі випадкового кроку [18]:

$$y(k) = a_0 + y(k-1) + \varepsilon(k). \quad (1.5)$$

У роботі [18] запропоновано використання моделі випадкового кроку з додатковою шумовою складовою та моделі випадкового кроку з шумом та дрейфом. У роботі [22] показано, як модель з трендом та сезонними компонентами може бути використаною як основа для процедури сезонного регулювання, практичні дослідження виконано для дослідження значень валового внутрішнього продукту США та значень кількості туристів у Іспанії.

Гетероскедастичними називають процеси зі змінною у часі дисперсією, а гомоскедастичними – із сталою [18]. Гетероскедастичні процеси є поширеними у випадку нестійкої економіки, наприклад, у перехідний період. Величина ризику, що стосується реалізації процесу, є поширеною економічною інтерпретацією дисперсії та стандартного відхилення [18]. Гетероскедастичні процеси докладно описано у роботі [23].

Для опису змінної дисперсії використовується модель авторегресії, побудована для квадратів оцінок залишків, що були отримані на попередньому етапі моделювання процесу [18]:

$$\varepsilon^2(k) = \alpha_0 + \alpha_1 \varepsilon^2(k-1) + \alpha_2 \varepsilon^2(k-2) + \dots + \alpha_q \varepsilon^2(k-q) + \nu(k), \quad (1.6)$$

де $\nu(k)$ – процес білого шуму.

Рівняння (1.6) називають авторегресійним умовно гетероскедастичним (АРУГ) [18].

АРУГ модель може бути розширеною за рахунок опису умовної дисперсії як процесу АРКС [18]. Нехай похибки описуються так [18]:

$$\varepsilon(k) = \nu(k)[h(k)]^{1/2},$$

де $\sigma_\nu^2 = 1$.

Тоді рівняння АРУГ моделі має такий вигляд [18]:

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (1.7)$$

Узагальнена модель АРУГ, яку називають УАРУГ(p,q), складається з двох компонентів – авторегресії та ковзного середнього відносно дисперсії процесу [18]. У роботі [18] показано, що для УАРУГ моделі збурення, що діє на процес $\{y(k)\}$, є процесом АРКС. Тому очікується, що залишки (похибки) моделі АРКС (попередньої моделі процесу) будуть відповідати за своїми характеристиками гетероскедастичному процесу [18]. Оскільки $E_{k-1}[\varepsilon(k)] = (h(k))^{1/2}$, то рівняння (1.7) можна записати так [18]:

$$E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (1.8)$$

Для прогнозування дисперсії використовують також експоненційну модель УАРУГ [18]:

$$\log(h^2(k)) = \alpha_0 + \sum_{i=1}^p \alpha_i \frac{|\varepsilon(k-i)|}{h(k-i)} + \sum_{i=1}^p \gamma_i \frac{\varepsilon(k-i)}{h(k-i)} + \sum_{i=1}^q \beta_i \log(h^2(k-i)). \quad (1.9)$$

Описані моделі гетероскедастичних процесів є відомими на сьогоднішній день, проте клас моделей гетероскедастичних процесів розширюється у зв'язку з необхідністю моделювання, прогнозування та керування ризиками.

1.2.2. Наявність сезонних ефектів

Сезонні ефекти, тобто вплив тієї чи іншої пори року на економічні процеси, можна пояснити так [24]: погода впливає на темпи будівництва – продуктивність праці максимальна у теплу пору року; у туризмі максимальну кількість відпочивальників (і туристів) на морях можна виявити у теплу пору року, а на гірськолижних курортах – взимку; розпродаж товарів перед різдвяними святами і Новим Роком, на День Подяки у США та на інші свята; розпродаж овочів, пік якого припадає на кінець літа і початок осені.

Сезонні ефекти можуть призвести до появи сезонної гетероскедастичності у часових рядах [24]. Прогнозування процесів без

врахування сезонних ефектів призводить до великих значень дисперсії похибок оцінок прогнозів і, відповідно, до їх неприйнятної якості [24].

Докладний аналіз сезонних ефектів та можливостей їх врахування у моделях наведено у роботі [19]. Зокрема, у роботі [19] запропоновано розраховувати сезонні різниці значень часового ряду для визначення сезонних ефектів та їх подальшого усунення.

Зазвичай сезонний ефект враховується включенням у модель змінної із запізненням, яке відповідає його періодичності [24]. Якщо сезонний ефект проявляється один раз на рік протягом одного кварталу, то під час обробки квартальних даних ($T_s = 3$ місяці) у модель потрібно ввести змінну із запізненням $d = 4$ [24].

У роботі [24] показано, що в адитивній формі можна враховувати сезонний ефект за допомогою основної (залежної) змінної або випадкової змінної у ковзному середньому, можливими є такі рівняння [24]:

$$y(k) = a_4 y(k-4) + \varepsilon(k), |a_4| < 1; \quad (1.10)$$

$$y(k) = \varepsilon(k) + \beta_4 \varepsilon(k-4). \quad (1.11)$$

Різноманітні підходи до побудови мультиплікативних сезонних моделей запропоновано у роботі [23]. Сезонні ефекти у мультиплікативній формі запропоновано описувати так [24]:

$$y(k)(1 - a_1 L) = \varepsilon(k)(1 + \beta_1 L)(1 + \beta_4 L^4).$$

Послідовність побудови математичної моделі часового ряду з сезонним ефектом наведено у роботі [24].

1.2.3. Методи відновлення пропущених значень часових рядів

Відновлення пропущених значень часових рядів є ефективнішим засобом вирішення проблеми пропущених даних, ніж виключення з дослідження часових рядів, що містять пропуски, тому що при цьому не відбувається втрата інформації.

Поширеним методом відновлення пропущених значень часових рядів є відновлення за допомогою середніх арифметичних. Заміна пропущених значень середніми арифметичними у випадку, коли є s пропущених значень, виконується таким чином [23]:

$$\hat{y}(k+i) = \frac{y(k) + y(k+s+1)}{2}, \quad (1.12)$$

де $i = 1, \dots, s$.

У випадку, якщо дані дозволяють це зробити, можна збільшувати ширину вікна для відновлення за допомогою середнього арифметичного з 2 значень до 4 або 6. Збільшувати ширину вікна до значень, що є більшими за 10, не рекомендується, оскільки значення ряду можуть суттєво відрізнятися на великих проміжках і це не призведе до покращення результатів відновлення.

Відновлення пропущених значень часового ряду може виконуватись також за допомогою методу експоненційного згладжування. Метод експоненційного згладжування має такий вигляд [25]:

$$S_k = \alpha \times y(k) + (1 - \alpha) \times S_{k-1}, \quad (1.13)$$

де S_k – значення експоненційного середнього в k -й момент часу; α – параметр згладжування; $y(k)$ – значення ряду в k -й момент часу; S_{k-1} – значення експоненційного середнього в $(k-1)$ -й момент часу.

В якості значення початкового наближення під час відновлення пропущених значень за допомогою методу експоненційного згладжування обирається останнє відоме значення перед пропущеним значенням ряду.

Для відновлення пропущених значень застосовується також метод авторегресії. Авторегресійна модель порядку p описується такою формулою:

$$y(k) = a_0 + \sum_{i=1}^p a_i \times y(k-i) + \varepsilon(k), E[\varepsilon(k)] = 0; \quad (1.14)$$

де $k = 1, \dots, N$ – це моменти часу, на основі даних про значення ряду у ці моменти часу будується математична модель; $y(k)$ – значення ряду в k -й момент часу; $y(k-i)$ – значення ряду в $(k-i)$ -й момент часу; a_0 , a_i – це

числові коефіцієнти авторегресійної моделі порядку p , де $i = 1, \dots, p$; $\varepsilon(k)$ – випадкове збурення [26].

Під час відновлення визначаються порядок та коефіцієнти авторегресійної моделі за відомими значеннями ряду, а потім вони використовуються для відновлення пропущених значень.

Наведені методи відновлення є поширеними, проте дослідження у цьому напрямі тривають та науковці створюють нові методи відновлення пропущених значень часових рядів.

1.3. Огляд методів ідентифікації моделей динаміки нестационарних процесів

Можливість знаходження істинних закономірностей часового ряду є важливою для подальшої обробки та прогнозування цього ряду. Особливо це важливо для специфічних даних, що містять великі шуми у даних про виміри, наприклад, для даних про сонячну активність, припливи та відпливи тощо. Тому дослідження методів ідентифікації істинної динаміки часових рядів є актуальним на сьогоднішній день.

1.3.1. Метод 13-місячного ковзного середнього

Метод 13-місячного ковзного середнього традиційно використовується для аналізу варіацій сонячної активності [27]. Метод 13-місячного ковзного середнього описується такою формулою [27]:

$$\hat{y}(k) = \frac{1}{24} \sum_{i=-6}^5 y(k+i) + \frac{1}{24} \sum_{i=-5}^6 y(k+i), \quad (1.15)$$

де $y(k)$ – місячне середнє значення для k -ого місяця; $\hat{y}(k)$ – 13-місячне ковзне середнє з центром у k -ому місяці.

Метод 13-місячного ковзного середнього є широко вживаним, проте цей метод може викривляти величини тривалості сонячного циклу, максимуму та мінімуму, що показано у роботі [27], оскільки він не відфільтровує високочастотні коливання з періодом менше одного року.

1.3.2. Модель сонячного циклу з використанням визначення параметрів сонячного циклу

У роботі [28] авторами запропоновано рівняння для опису сонячного циклу, що також можуть використовуватись для визначення істинних закономірностей зміни значень сонячного циклу, хоча вони і не набули такої популярності, як метод 13-місячного ковзного середнього, що описується формулою (1.15). Метод, запропонований у роботі [28], описується такими формулами:

$$f(i) = \frac{a \times (i - i_0)^3}{\exp((i - i_0)^2 / b^2) - c}, \quad (1.16)$$

$$b(a) = 27,12 + 25,15 / (a \times 10^3)^{1/4}, \quad (1.17)$$

$$a(n) = 9,2 \times 10^{-4} + 8,0 \times 10^{-5} n, \quad (1.18)$$

де $f(i)$ – функція, що описує фази росту і падіння сонячних даних; a – амплітуда циклу; i – момент часу, що вимірюється у місяцях; b – час у місяцях від мінімуму до максимуму; c – асиметрія циклу; i_0 – початковий момент часу; n – номер циклу.

Рівняння (1.16) показує фази зростання та спадання у сонячному циклі та згладжені переходи між значеннями поблизу мінімуму та максимуму [28]. Це рівняння є схожим на рівняння Планка, але містить чотири вільних параметри та має більш швидке спадання значень після максимуму [28].

1.3.3. Використання різницевого рівняння для опису статистичних даних

Різницеві рівняння широко застосовуються завдяки простоті структури та наявності надійних методів оцінювання під час моделювання економічних, фінансових, екологічних та інших процесів. Простим прикладом різницевого рівняння є стохастичне рівняння авторегресії першого порядку з одиничним коефіцієнтом (окремий випадок) і затриманим у часі значенням основної змінної:

$$y(k) = y(k-1) + \varepsilon(k), \quad (1.19)$$

де $y(k)$ – основна змінна; $\varepsilon(k)$ – випадкова величина, яка відображає вплив різноманітних невимірюваних факторів на основну змінну [24]. Насамперед, це випадкові збурення, що діють на процес [24]. Найчастіше припускають, що випадкова величина має нормальний розподіл [24].

Рівняння (1.19) називають ще рівнянням, яке описує процес випадкового кроку (випадкове блукання) [24]. Рівняння (1.19) може бути записане також у формі першої різниці:

$$\Delta y(k) = \varepsilon(k),$$

де $\Delta y(k) = y(k) - y(k-1)$ [24]. Саме можливість використання різниць перших і вищих порядків для опису часових рядів зумовило використання назви «різницеві рівняння» [24].

Різницеві рівняння, у правій частині яких наявні минулі (затримані) виміри основної змінної, називають авторегресійними (АР), тобто регресія змінної спрямована на саму себе [24]. Рівняння авторегресії p -го порядку описується формулою (1.14).

Якщо для модельованого процесу можна виявити вхідну змінну, то її записують у правій частині [24]. Якщо ця змінна є випадковим процесом, то таке рівняння називають авторегресією з ковзним середнім (АРКС) [24].

Авторегресійна модель порядку p з ковзним середнім порядку q описується такою формулою [26]:

$$y(k) = a_0 + \sum_{i=1}^p a_i \times y(k-i) + v(k) + \sum_{j=1}^q b_j \times v(k-j), \quad (1.20)$$

де $k=1, \dots, N$ – це моменти часу, на основі даних про значення ряду у ці моменти часу будується математична модель; $y(k)$ – значення ряду в k -й момент часу; $y(k-i)$ – значення ряду в $(k-i)$ -й момент часу; a_0, a_i – це числові коефіцієнти авторегресійної моделі порядку p , де $i=1, \dots, p$; $v(k)$ – значення залишків моделі низького порядку в k -й момент часу; $v(k-j)$ – значення залишків моделі низького порядку в $(k-j)$ -й момент часу; b_j – числові коефіцієнти, обчислені за авторегресійною моделлю порядку p з ковзним середнім порядку q , де $j=1, \dots, q$.

У випадку ковзного середнього повинна виконуватись така умова [24]:

$$\sum_{j=1}^q b_j = 1.$$

Характеристичне рівняння, записане для рівняння (1.20), може мати одиничні корені [24]. Такі процеси називають процесами АРІКС(p, d, q), де d – кількість одиничних коренів характеристичного рівняння [24]. Процеси цього класу є нестационарними, тобто вони мають тренд, порядок якого визначається кількістю одиничних коренів [24].

1.4. Методи і моделі для прогнозування часових рядів

Існує багато методів для прогнозування часових рядів на сьогоднішній день. Поширеними серед них є регресійні методи, методи експоненційного згладжування, фільтр Калмана, методи групового урахування аргументів. Вдосконалення цих методів залишається актуальним, тому що отримання

точніших оцінок прогнозних значень часових рядів є важливим для дослідження фінансових, економічних, екологічних та інших процесів.

1.4.1. Регресійні моделі

Рівняння авторегресії у загальному випадку описує вплив значень попередніх станів на поточний стан процесу. Регресійні моделі широко використовуються для прогнозування часових рядів.

Найпоширеніша авторегресійна модель порядку p описується формулою (1.14). Застосовується для прогнозування часових рядів також парна регресія, що містить у правій частині незалежну змінну (регресор):

$$y(k) = a_0 + a_1 x(k) + \varepsilon(k),$$

де $x(k)$ – регресор (незалежна або екзогенна змінна); $y(k)$ – залежна змінна (ендогенна змінна) [24].

Множинна регресія відображає вплив декількох незалежних змінних на залежну [24]:

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + \dots + a_p x_p(k) + \varepsilon(k),$$

де $x_1(k), \dots, x_p(k)$ – регресори рівняння. Таке рівняння може містити також авторегресійну частину.

Змішана регресія описується таким рівнянням (поєднує авторегресію та множинну регресію) [24]:

$$y(k) = a_0 + \sum_{i=1}^l a_i y(k-i) + b_1 x_1(k) + b_2 x_2(k) + \dots + b_p x_p(k) + \varepsilon(k).$$

Авторегресія з ковзним середнім (АРКС(p,q)) описується рівнянням (1.20) і також використовується для прогнозування значень часових рядів. Докладний огляд нелінійних регресійних моделей та підходів до роботи з ними виконано у роботі [29].

Для прогнозування використовуються також моделі гетероскедастичних процесів, наприклад, поширені моделі гетероскедастичних процесів описуються рівняннями (1.7), (1.8), (1.9).

Таким чином, на сьогоднішній день регресійні та авторегресійні моделі є поширеними, різні їх види використовуються для прогнозування часових рядів.

1.4.2. Методи експоненційного згладжування

Метод експоненційного згладжування є дуже поширеним. Зокрема, він застосовується у випадках, коли потрібно працювати в умовах відсутності повної та надійної інформації про закономірності розвитку процесу. Метод експоненційного згладжування був вперше запропонований Брауном [30] і згодом розширений Холтом [31]. Цей метод описується формулою (1.13).

Гарднер у своїй роботі [32] виконує аналіз теоретичних та емпіричних аргументів на користь вибору параметра згладжування, також він розглянув різноманітні моделі експоненційного згладжування у термінах сезонності (відсутня, адитивна сезонність, мультиплікативна) і тренду (відсутній, лінійний тренд, експоненційний, демпфований).

Також розповсюдженим є подвійне експоненційне згладжування. Метод подвійного експоненційного згладжування описується так [33]:

$$S_k = \alpha \times y(k) + (1 - \alpha) \times (S_{k-1} - b_{k-1}),$$

$$b_k = \beta \times (S_k - S_{k-1}) + (1 - \beta) \times b_{k-1},$$

де α і β – параметри згладжування $0 < \alpha < 1$, $0 < \beta < 1$.

Гарднер у роботі [33] описує останні дослідження стосовно експоненційного згладжування, зокрема, він називає повне статистичне обґрунтування експоненційного згладжування на основі нового класу моделей, що мають одне джерело похибок, найважливішим теоретичним досягненням. Порівняльний аналіз останніх методів експоненційного згладжування з зазначенням наявності сезонності та тренду наведено на рис. 1.1 [33].

Trend	Seasonality		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$ $\hat{X}_t(m) = S_t$	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)S_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)S_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t I_{t-p+m}$
	$S_t = S_{t-1} + \alpha e_t$ $\hat{X}_t(m) = S_t$	$S_t = S_{t-1} + \alpha e_t$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = S_{t-1} + \alpha e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = S_t I_{t-p+m}$
A (Additive)	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha \gamma e_t$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha \gamma e_t$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = S_{t-1} + T_{t-1} + \alpha e_t/I_{t-p}$ $T_t = T_{t-1} + \alpha \gamma e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
DA (Damped additive)	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)\phi T_{t-1}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i$	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)\phi T_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)\phi T_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = \left(S_t + \sum_{i=1}^m \phi^i T_i \right) I_{t-p+m}$
	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t$ $T_t = \phi T_{t-1} + \alpha \gamma e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i$	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t$ $T_t = \phi T_{t-1} + \alpha \gamma e_t$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^m \phi^i T_i + I_{t-p+m}$	$S_t = S_{t-1} + \phi T_{t-1} + \alpha e_t/I_{t-p}$ $T_t = \phi T_{t-1} + \alpha \gamma e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = \left(S_t + \sum_{i=1}^m \phi^i T_i \right) I_{t-p+m}$
M (Multiplicative)	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} R_{t-1})$ $R_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)R_{t-1}$ $\hat{X}_t(m) = S_t R_t^m$	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)S_{t-1} R_{t-1}$ $R_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)R_{t-1}$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t R_t^m + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)S_{t-1} R_{t-1}$ $R_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)R_{t-1}$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = (S_t R_t^m) I_{t-p+m}$
	$S_t = S_{t-1} R_{t-1} + \alpha e_t$ $R_t = R_{t-1} + \alpha \gamma e_t/S_{t-1}$ $\hat{X}_t(m) = S_t R_t^m$	$S_t = S_{t-1} R_{t-1} + \alpha e_t$ $R_t = R_{t-1} + \alpha \gamma e_t/S_{t-1}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t R_t^m + I_{t-p+m}$	$S_t = S_{t-1} R_{t-1} + \alpha e_t/I_{t-p}$ $R_t = R_{t-1} + \alpha \gamma e_t/S_{t-1} I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t R_t^m) I_{t-p+m}$
DM (Damped multiplicative)	$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} R_{t-1}^\phi)$ $R_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)R_{t-1}^\phi$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^m \phi^i}$	$S_t = \alpha(X_t - I_{t-p}) + (1 - \alpha)S_{t-1} R_{t-1}^\phi$ $R_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)R_{t-1}^\phi$ $I_t = \delta(X_t - S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^m \phi^i} + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1 - \alpha)(S_{t-1} R_{t-1}^\phi)$ $R_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)R_{t-1}^\phi$ $I_t = \delta(X_t/S_t) + (1 - \delta)I_{t-p}$ $\hat{X}_t(m) = \left(S_t R_t^{\sum_{i=1}^m \phi^i} \right) I_{t-p+m}$
	$S_t = S_{t-1} R_{t-1}^\phi + \alpha e_t$ $R_t = R_{t-1}^\phi + \alpha \gamma e_t/S_{t-1}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^m \phi^i}$	$S_t = S_{t-1} R_{t-1}^\phi + \alpha e_t$ $R_t = R_{t-1}^\phi + \alpha \gamma e_t/S_{t-1}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^m \phi^i} + I_{t-p+m}$	$S_t = S_{t-1} R_{t-1}^\phi + \alpha e_t/I_{t-p}$ $R_t = R_{t-1}^\phi + \alpha \gamma e_t/S_{t-1} I_{t-p}$ $I_t = I_{t-p} + \delta(1 - \alpha)e_t/S_t$ $\hat{X}_t(m) = \left(S_t R_t^{\sum_{i=1}^m \phi^i} \right) I_{t-p+m}$

Рис. 1.1. Порівняльний аналіз методів експоненційного згладжування з зазначенням наявності сезонності та тренду

Наведені на рис. 1.1 методи експоненційного згладжування є найбільш відомими на даний час. Проте дослідження стосовно застосування експоненційного згладжування для розв'язування прикладних задач тривають, тому продовжується створення нових методів.

1.4.3. Фільтр Калмана

Поширеним є дискретне представлення фільтра Калмана. Розглядається лінійна динамічна система, що описується такими рівняннями [34]:

$$\mathbf{X}_{k+1} = \mathbf{\Phi}_{k+1,k} \times \mathbf{X}_k + \mathbf{w}_k, \quad (1.21)$$

$$\mathbf{z}_k = \mathbf{H}_k \times \mathbf{X}_k + \mathbf{V}_k, \quad (1.22)$$

$$E[\mathbf{V}_k] = 0, \quad (1.23)$$

$$E[\mathbf{V}_k \times \mathbf{V}_i^T] = \mathbf{R}_k \times \delta(k-i), \quad (1.24)$$

$$E[\mathbf{w}_k] = \mathbf{q}_k, \quad (1.25)$$

$$E[(\mathbf{w}_i - \mathbf{q}_i) \times (\mathbf{w}_k - \mathbf{q}_k)^T] = \mathbf{Q}_k \times \delta(i-k), \quad (1.26)$$

$$E[\mathbf{V}_i \times \mathbf{w}_k^T] = 0, \quad (1.27)$$

де \mathbf{X} – n -вимірний вектор стану системи; $\mathbf{\Phi}$ – матриця динаміки розмірністю $n \times n$; \mathbf{w} – n -вимірний вектор вхідного шуму; \mathbf{z} – p -вимірний вектор вимірів; \mathbf{H} – матриця спостережень розмірністю $p \times n$; \mathbf{V} – p -вимірний вектор шуму вимірів; $\delta(i-k)$ – дельта-функція Дірака.

Оптимальний дискретний фільтр Калмана описується такими рівняннями [35]:

екстраполяція

$$\hat{\mathbf{X}}_{k+1,k} = \mathbf{\Phi}_{k+1,k} \times \hat{\mathbf{X}}_{k,k} + \mathbf{q}_k, \quad (1.28)$$

$$\mathbf{P}_{k+1,k} = \mathbf{\Phi}_{k+1,k} \times \mathbf{P}_{k,k} \times \mathbf{\Phi}_{k+1,k}^T + \mathbf{Q}_k; \quad (1.29)$$

фільтрація

$$\hat{\mathbf{X}}_{k,k} = \hat{\mathbf{X}}_{k,k-1} + \mathbf{K}_k \times \mathbf{v}_k, \quad (1.30)$$

$$\mathbf{K}_k = \mathbf{P}_{k,k-1} \times \mathbf{H}_k^T \times (\mathbf{H}_k \times \mathbf{P}_{k,k-1} \times \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (1.31)$$

$$\mathbf{P}_{k,k} = (\mathbf{I} - \mathbf{K}_k \times \mathbf{H}_k) \times \mathbf{P}_{k,k-1}, \quad (1.32)$$

де $\mathbf{v}_k = \mathbf{z}_k - \mathbf{H}_k \times \hat{\mathbf{X}}_{k,k-1}$ – p -вимірний вектор нев'язки фільтра; \mathbf{K} – матриця коефіцієнтів підсилення фільтра Калмана розмірністю $n \times p$; \mathbf{P} – коваріаційна

матриця похибок фільтрації розмірністю $n \times n$; \mathbf{H} – матриця спостережень розмірністю $p \times n$; \mathbf{R} – матриця дисперсій шуму вимірів розмірністю $p \times p$; \mathbf{q} – n -вимірний вектор математичних сподівань шуму стану; \mathbf{Q} – матриця дисперсій шуму стану розмірністю $n \times n$.

У роботі [34] отримано явний вигляд основних характеристик дискретного фільтра Калмана. Для рівняння стану виконуються такі співвідношення [34]:

$$E[\mathbf{X}_0] = \mathbf{M}_0, \quad (1.33)$$

$$E[(\mathbf{X}_0 - \mathbf{M}_0) \times (\mathbf{X}_0 - \mathbf{M}_0)^T] = \mathbf{P}_0. \quad (1.34)$$

Оцінка вектора стану у роботі [34] знаходиться за такими формулами:

$$\hat{\mathbf{X}}_{k,k} = \Psi_{k,0} \times \hat{\mathbf{X}}_{0,0} + \sum_{i=1}^k \Psi_{k,i} \times \mathbf{K}_i \times \mathbf{z}_i, \quad (1.35)$$

$$\mathbf{K}_i = \mathbf{P}_{i,i} \times \mathbf{H}_i^T \times \mathbf{R}_i^{-1}, \quad (1.36)$$

$$\hat{\mathbf{X}}_{0,0} = \mathbf{M}_0, \quad (1.37)$$

$$\mathbf{P}_{0,0} = \mathbf{P}_0, \quad (1.38)$$

де Ψ – перехідна матриця фільтра розмірністю $n \times n$.

Перехідна матриця фільтра визначається такими рівняннями у роботі [34]:

$$\Psi_{k,i} = \mathbf{P}_{k,k} \times \Phi_{j,k}^T \times \mathbf{P}_{j,j}^{-1}, \text{ при } \mathbf{P}_{0,0} > 0, \quad (1.39)$$

$$\Psi_{k,j} = \Phi_{k,j} \times (\mathbf{I} - \mathbf{P}_{j,j} \times \sum_{i=j+1}^k \Phi_{i,j}^T \times \mathbf{H}_i^T \times \mathbf{R}_i^{-1} \times \mathbf{H}_i \times \Phi_{i,j})^{-1}, \text{ при } \mathbf{P}_{0,0} \geq 0. \quad (1.40)$$

Коваріаційна матриця похибок фільтрації визначається такими рівняннями у роботі [34]:

$$\mathbf{P}_{k,k} = (\sum_{i=1}^k \Phi_{i,k}^T \times \mathbf{H}_i^T \times \mathbf{R}_i^{-1} \times \mathbf{H}_i \times \Phi_{i,k} + \Phi_{0,k}^T \times \mathbf{P}_{0,0}^{-1} \times \Phi_{0,k})^{-1}, \mathbf{P}_{0,0} > 0, \quad (1.41)$$

$$\mathbf{P}_{k,k} = \Phi_{k,0} (\mathbf{I} + \mathbf{P}_{0,0} \times \sum_{i=1}^k \Phi_{i,0}^T \times \mathbf{H}_i^T \times \mathbf{R}_i^{-1} \times \mathbf{H}_i \times \Phi_{i,0})^{-1} \mathbf{P}_{0,0} \Phi_{k,0}^T, \mathbf{P}_{0,0} \geq 0. \quad (1.42)$$

Критерієм оптимальності фільтра Калмана є мінімізація математичного сподівання суми квадратів похибок оцінки вектора стану [34].

Розповсюдженим є також субоптимальний фільтр Калмана. Але при розв'язанні задач ідентифікації субоптимальний фільтр Калмана, що будується для системи з неточно заданими параметрами, може бути розглянутим лише як допоміжний фільтр, що використовується для формування послідовності нев'язок. Тому під час вибору початкових значень невідомих параметрів при побудові такого фільтра більш суттєвими можуть бути вимоги до його структурної простоти. Вимогам до структурної простоти, що вказані, відповідає фільтр Калмана, що є оптимальним для моделі динамічної системи, яка є вільною від шуму стану [34]. У роботі [34] показано, що за допомогою такого допоміжного фільтра аналітично можна представити статистичні характеристики нев'язок фільтра та похибок фільтрації як явні функції невідомих середніх значень і коваріаційних матриць шумів стану та вимірів.

Фактична модель описується такими рівняннями [34]:

$$\mathbf{X}_i = \Phi \mathbf{X}_{i-1} + \mathbf{G} \mathbf{W}_{i-1}, \quad (1.43)$$

$$\mathbf{Z}_i = \mathbf{H} \mathbf{X}_i + \mathbf{V}_i, \quad (1.44)$$

$$E[\mathbf{W}_i] = \mathbf{q}^*, \quad (1.45)$$

$$E[(\mathbf{W}_i - \mathbf{q}^*)(\mathbf{W}_k - \mathbf{q}^*)^T] = \mathbf{Q}^* \delta(i - k), \quad (1.46)$$

$$E[\mathbf{V}_i] = \mathbf{r}^*, \quad (1.47)$$

$$E[(\mathbf{V}_i - \mathbf{r}^*)(\mathbf{V}_k - \mathbf{r}^*)^T] = \mathbf{R}^* \delta(i - k), \quad (1.48)$$

де \mathbf{X} – n -вимірний вектор стану; \mathbf{Z} – p -вимірний вектор вимірів.

Розрахункова модель описується такими рівняннями [34]:

$$\mathbf{X}_i = \Phi \mathbf{X}_{i-1}, \quad (1.49)$$

$$\mathbf{Z}_i = \mathbf{H} \mathbf{X}_i + \mathbf{V}_i, \quad (1.50)$$

$$E[\mathbf{V}_i] = \mathbf{0}, \quad (1.51)$$

$$E[\mathbf{V}_i \mathbf{V}_k^T] = \mathbf{R} \delta(i - k). \quad (1.52)$$

Нев'язки субоптимального фільтра з фіксованою пам'яттю n формуються відповідно до такого рівняння [34]:

$$\mathbf{v}_i^{(n)} = \mathbf{Z}_i - \mathbf{H}\mathbf{\Phi} \hat{\mathbf{X}}_{i-1,i-1}^{(n)}, \quad (1.53)$$

де

$$\hat{\mathbf{X}}_{i,i}^{(n)} = \begin{cases} \mathbf{P}_{n,n} \sum_{j=1}^n (\mathbf{\Phi}^{(j-n)})^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{Z}_j, & i = n, \\ \mathbf{\Phi} \hat{\mathbf{X}}_{i-1,i-1}^{(n)} + \mathbf{K} \mathbf{v}_i^{(n)}, & i > n, \end{cases} \quad (1.54)$$

$$\mathbf{K} = \mathbf{P}_{n,n} \mathbf{H}^T \mathbf{R}^{-1}, \quad (1.55)$$

$$\mathbf{P}_{n,n} = \left(\sum_{j=1}^n (\mathbf{\Phi}^{(j-n)})^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{\Phi}^{(j-n)} \right)^{-1}. \quad (1.56)$$

Нев'язки субоптимального фільтра з фіксованою пам'яттю $(n+s)$ формуються відповідно до такого рівняння [34]:

$$\mathbf{v}_i^{(n+s)} = \mathbf{Z}_i - \mathbf{H}\mathbf{\Phi} \hat{\mathbf{X}}_{i-1,i-1}^{(n+s)}, \quad (1.57)$$

де

$$\hat{\mathbf{X}}_{i,i}^{(n+s)} = \begin{cases} \hat{\mathbf{X}}_{i,i}^{(n)}, & s = 0, \\ \mathbf{\Phi} \hat{\mathbf{X}}_{i-1,i-1}^{(n+s-1)} + \mathbf{K}_s \mathbf{v}_i^{(n+s-1)}, & s > 0, \end{cases} \quad (1.58)$$

$$\mathbf{K}_s = \mathbf{P}_{n+s,n+s} \mathbf{H}^T \mathbf{R}^{-1}. \quad (1.59)$$

Середнє значення нев'язок описується так [34]:

$$E[\mathbf{v}_i^{(n)}] = \sum_{j=1}^n \mathbf{\Omega}_{n,j} \mathbf{q}^* - \sum_{j=1}^n \mathbf{S}_{n,j} \mathbf{r}^* + \mathbf{r}^*, \quad (1.60)$$

де складові описуються такими рівняннями [34]:

$$\mathbf{\Omega}_{n,j} = \mathbf{H}\mathbf{\Phi} \mathbf{P}_{n,n} (\mathbf{\Phi}^{(j-n)})^T \mathbf{P}_{j,j}^{-1} \mathbf{G}, \quad (1.61)$$

$$\mathbf{S}_{n,j} = \mathbf{H}\mathbf{\Phi} \mathbf{P}_{n,n} (\mathbf{\Phi}^{(j-n)})^T \mathbf{H}^T \mathbf{R}^{-1}. \quad (1.62)$$

Коваріація нев'язок описується так [34]:

$$\mathbf{C}_0 = \sum_{j=1}^n \mathbf{\Omega}_{n,j} \mathbf{Q}^* \mathbf{\Omega}_{n,j}^T + \sum_{j=1}^n \mathbf{S}_{n,j} \mathbf{R}^* \mathbf{S}_{n,j}^T + \mathbf{R}^*. \quad (1.63)$$

Кореляція нев'язок при $(i > k)$ описується такими рівняннями [34]:

$$\mathbf{C}_{i-k} = \sum_{j=1}^n \mathbf{\Omega}_{n+i-k,j} \mathbf{Q}^* \mathbf{\Omega}_{n,j}^T + \sum_{j=1}^n \mathbf{S}_{n+i-k,j} \mathbf{R}^* \mathbf{S}_{n,j}^T - \mathbf{S}_{n+i-k,n+1} \mathbf{R}^*, \quad (1.64)$$

$$\mathbf{C}_{i-k} = \sum_{j=1}^n \mathbf{\Omega}_{n+i-k,j} \mathbf{Q}^* \mathbf{\Omega}_{n,j}^T + \sum_{j=1}^n \mathbf{S}_{n+i-k,j} \mathbf{R}^* \mathbf{S}_{n,j}^T - \mathbf{S}_{n+i-k,n+1} \mathbf{R}^*, \mathbf{R}^* = c \mathbf{R}. \quad (1.65)$$

Таким чином, рівняння (1.43)–(1.65) повністю описують субоптимальний фільтр Калмана у загальному вигляді.

1.4.4. Метод групового урахування аргументів

Академік О.Г. Івахненко запропонував метод групового урахування аргументів (МГУА) наприкінці 60-х – 70-х років 20 сторіччя [36-39]. У роботі [40] показано, що у цьому методі використовуються ідеї самоорганізації і механізми живої природи – схрещування (гібридизація) і селекція (добір), схематичну постановку задачі для розв'язання її за допомогою МГУА показано на рис. 1.2 [40].

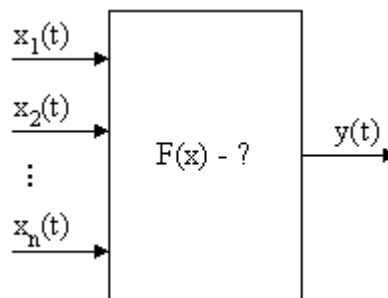


Рис. 1.2. Схематична постановка задачі для розв'язання її за допомогою МГУА

Нехай є вибірка з N спостережень вхідних векторів $X(i)$ та вихідних $Y(i)$ [40]:

$$\begin{array}{c} \{X(1) \quad Y(1)\} \\ \{X(2) \quad Y(2)\} \\ \dots\dots\dots \\ \{X(N) \quad Y(N)\} \end{array}$$

За результатами спостережень треба визначити $F(x)$, причому навіть структура моделі $F(x)$ невідома [40].

Найбільш повна залежність між входами $X(i)$ і виходами $Y(i)$ описується узагальненим поліномом Колмогорова-Габора [40].

Нехай є $X = \{x_1, \dots, x_N\}$, тоді такий поліном має вигляд [40]:

$$Y = a_0 + \sum_{i=1}^N a_i x_i + \sum_{j=1}^N \sum_{i \leq j} a_{ij} x_i x_j + \sum_{i=1}^N \sum_{j \leq i} \sum_{k \leq j} a_{ijk} x_i x_j x_k + \dots,$$

де всі коефіцієнти a_i не відомі.

При визначенні значень коефіцієнтів при побудові моделі використовується критерій регулярності (точності) в якості критерію [40]:

$$\overline{\varepsilon^2} = \frac{1}{N} \cdot \sum_{i=1}^N (y_i - f(x_i))^2.$$

При цьому потрібно, щоб виконувалась така умова [40]: $\overline{\varepsilon^2} \rightarrow \min$.

Докладний огляд методу групового урахування аргументів виконано також у роботі [41]. Подальші дослідження виконано у роботі [42]. Існує два підходи для вибору часткових описів і побудови МГУА [40]: точнісний та робастний.

Багаторядні алгоритми МГУА застосовуються для розв'язання некоректних чи недокінця визначених задач моделювання [40]. Методи регресійного аналізу у цьому випадку не можна застосувати, тому що вони не дозволяють побудувати єдину модель, що за межами інтервалу інтерполяції є адекватною процесу [40]. Також у роботі [40] показано, що багаторядні алгоритми МГУА успішно застосовуються у випадках, коли вихідних даних достатньо для застосування однорядних МГУА, але за таких умов багаторядні часто показують кращі результати.

Опис алгоритму багаторядного поліноміального МГУА

Перша ітерація

Крок 1. З множини виходів $X = \{X_1, X_2, \dots, X_n\}$ обираються пари аргументів X_i, X_j і складаються часткові описи такого вигляду [40]:

$$Y_k^{(1)} = \varphi(X_i, X_j), i \neq j, i, j = \overline{1, N}, \quad (1.66)$$

при цьому використовують часткові описи квадратичного типу:

$$Y_k^{(1)} = a_0 + a_i X_i + a_j X_j + a_{ij} X_i X_j + a_{ii} X_i^2 + a_{jj} X_j^2. \quad (1.67)$$

Число часткових описів 1-го ряду дорівнює $M = n(n-1)/2$ [40].

Крок 2. За навчальною вибіркою з використанням методу найменших квадратів (МНК) для кожного опису знаходяться оцінки невідомих коефіцієнтів $\bar{a}_0, \bar{a}_i, \bar{a}_j, \bar{a}_{ij}, \bar{a}_{ii}, \bar{a}_{jj}$ [40].

Крок 3. За критерієм мінімуму $\bar{\varepsilon}^2$ на перевірочній послідовності відбирається F_1 кращих моделей, тобто реалізують процедуру селекції [40]. Величина F_1 називається свободою вибору, при цьому $F_1 < M$ [40]. Виходи цих моделей слугують аргументами-входами для конструювання моделей другого ряду [40].

Крок 4. Знаходиться $\bar{\varepsilon}^2(0) = \min_k \bar{\varepsilon}_k^2(0)$ [40].

m-а ітерація

Крок 1. Формуються часткові описи такого вигляду [40]:

$$Y_k^{(m)} = a_0^{(m)} + a_i^{(m)} X_i + a_j^{(m)} X_j + a_{ij}^{(m)} X_i X_j + a_{ii}^{(m)} X_i^2 + a_{jj}^{(m)} X_j^2, \quad (1.68)$$

$$k = 1, \dots, \frac{F_1(F_1 - 1)}{2}.$$

Крок 2. З використанням МНК для кожного опису розраховуються оцінки $\bar{a}_0^{(m)}, \bar{a}_i^{(m)}, \bar{a}_j^{(m)}, \bar{a}_{ij}^{(m)}, \bar{a}_{ii}^{(m)}, \bar{a}_{jj}^{(m)}$ [40].

Крок 3. За перевіркою послідовністю знаходиться для кожного часткового опису величина критерію

$$\bar{\varepsilon}_k^2(m) = \frac{1}{N_{nep}} \sum_{i=1}^{N_{nep}} (Y_i - Y_{ki}^{(m)})^2, \quad (1.69)$$

де N_{nep} – обсяг перевірконої вибірки [40].

Крок 4. Знаходиться $\bar{\varepsilon}^2(m) = \min_k \bar{\varepsilon}_k^2(m)$ [40]. Перевіряється умова $\bar{\varepsilon}^2(m) > \bar{\varepsilon}^2(m-1)$, де $\bar{\varepsilon}^2(m), \bar{\varepsilon}^2(m-1)$ – величини критерію точності для найкращих моделей $(m-1)$ -го і m -го ряду селекції відповідно [40]. Якщо так, то кінець [40]. Шукана модель вибирається з часткових описів m^{-1} -го рівня, на якому досягається мінімальна помилка $\bar{\varepsilon}^2(m-1)$ [40]. Інакше перехід до конструювання наступного ряду часткових описів [40]. При цьому проводиться добір (селекція) F_2 кращих описів [40].

Заключний етап.

У роботі [40] показано, що на заключному етапі у початковому просторі станів розраховуються вирази для моделі, що будується, за рахунок руху від кінця до початку і виконання послідовної заміни змінних [40].

Поширеними також є комбінаторний алгоритм МГУА, алгоритм МГУА з послідовним виділенням трендів та нечіткий МГУА.

1.5. Застосування систем підтримки прийняття рішень для аналізу та прогнозування часових рядів

Створення систем підтримки прийняття рішень для аналізу та прогнозування часових рядів є актуальною задачею на сьогоднішній день, тому що створення таких систем дозволяє проводити швидко та автоматизовану обробку часових рядів.

Детальний алгоритм побудови системи підтримки прийняття рішень для аналізу та прогнозування часових рядів наведено у роботі [43]:

Крок 1 [43]: на цьому кроці виконується аналіз та попередня обробка даних: якщо є пропуски, то виконується їх заповнення; логарифмування; згладжування

екстремальних значень часового ряду; у діапазоні від -1 до $+1$ виконується нормування; диференціювання (обчислення різниць першого та вищих порядків); цифрова фільтрація; пряме і зворотне перетворення Фур'є; бутстреп (розмноження вибірки даних). Виконання переходу на крок 2.

Крок 2 [43]: з використанням кореляційних функцій вищих порядків або тесту Фішера виконується перевірка на наявність нелінійностей. Виконується перехід до кроку 3, якщо процес містить нелінійність, інакше – перехід до кроку 4.

Крок 3 [43]: визначається порядок нелінійностей, будується модель за МГУА та виконується лінійна апроксимація процесу для того, щоб порівняти її з нелінійною моделлю. Виконання переходу на крок 9.

Крок 4 [43]: за тестом Дікі-Фуллера процес перевіряється на стаціонарність. Перехід до кроку 9, якщо процес стаціонарний, інакше – перехід до кроку 5.

Крок 5 [43]: виконується перевірка на наявність гетероскедастичності за допомогою тестів Голдфельда-Квандта, Уайта, Бройша-Пагана/Годфрі. З використанням тесту Дікі-Фуллера перевіряється наявність тренду. За методикою Йохансена або Інгла-Грейнджера проводиться аналіз коінтегрованості процесів. Перехід до кроку 6, якщо процес гетероскедастичний; перехід до кроку 7, якщо процес містить тренд; інакше перехід до кроку 8.

Крок 6 [43]: для опису гетероскедастичності визначається тип моделі: УАРУГ(p, q), експоненційна УАРУГ(p, q), УАРУГ-М або інша. Виконується перехід до кроку 9 після вибору адекватної моделі.

Крок 7 [43]: для моделювання або вилучення тренду визначається можливість опису тренду поліноміальною, логарифмічною або іншими функціями та визначається порядок інтегрованості процесу. Перехід до кроку 9 після вилучення тренду.

Крок 8 [43]: виконується побудова моделі корекції похибок для коінтегрованих процесів. Перехід до кроку 10.

Крок 9 [43]: виконується побудова моделі часового ряду, розраховуються критерії адекватності цієї моделі. Виконується перехід до кроку 10.

Крок 10 [43]: побудова функції прогнозування на основі обраної моделі. Розраховуються прогнозні оцінки та визначаються оцінки їх точності.

Крок 11 [43]: якщо особу, що приймає рішення, не задовольняє точність прогнозу, то виконується уточнення вхідних даних і перехід до кроку 1. В іншому випадку процес закінчується.

Схему процесу прогнозування часових рядів наведено на рис. 1.3 [43].

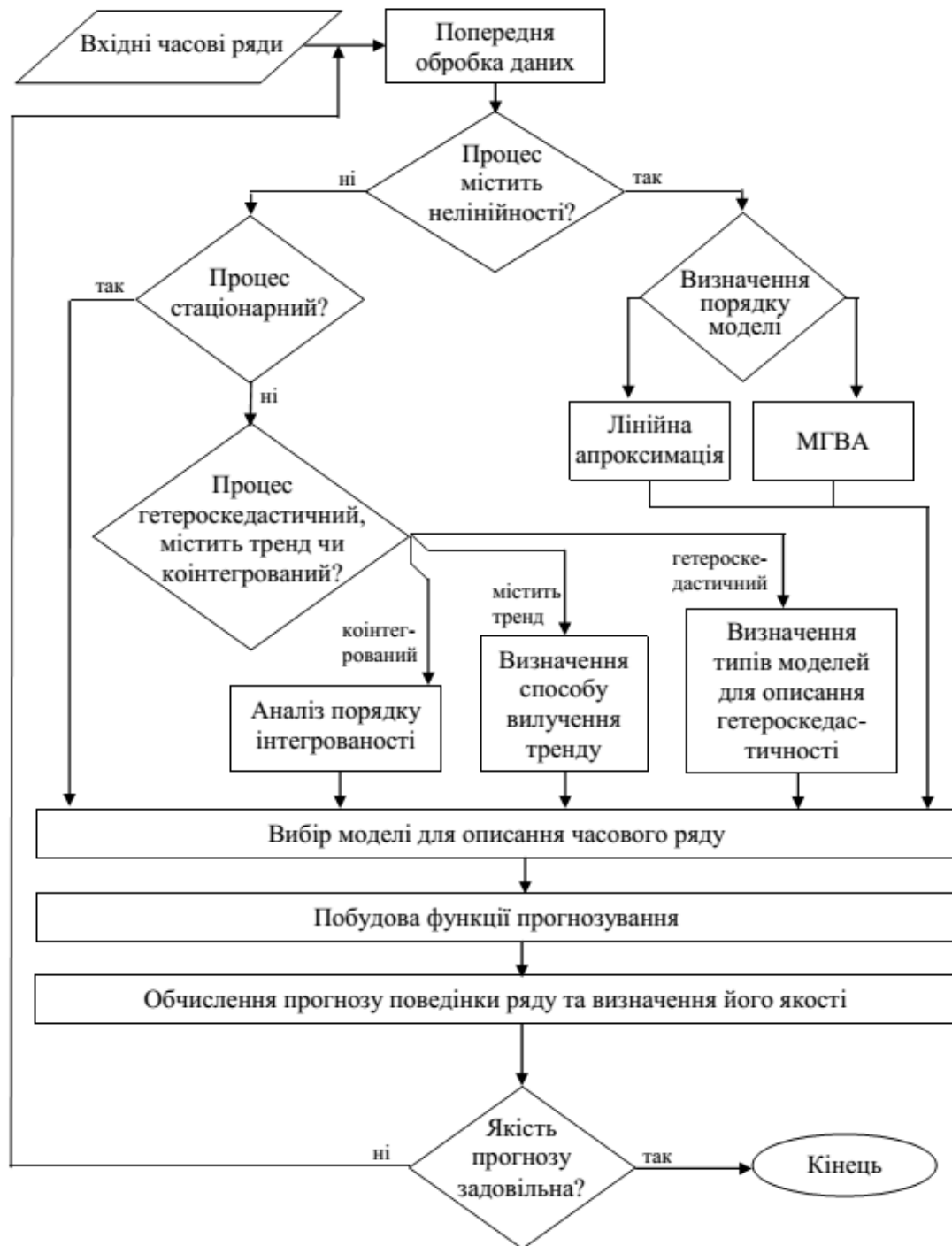


Рис. 1.3. Схема процесу прогнозування часових рядів

Дослідження стосовно розробки системи підтримки прийняття рішень для прогнозування динаміки стаціонарних та нестаціонарних процесів різної природи наведено також у роботі [44]. У роботі [45] запропоновано інформаційну систему, що забезпечує виконання таких задач: фрактальний аналіз часового ряду перед прогнозуванням, реалізація комбінованих моделей прогнозування рівнів та знаків приростів часових рядів, ідентифікація моментів зміни їх тенденцій для прийняття фінансових рішень. У роботі [46] наведено результати наукових досліджень та практичних розробок методів експертного оцінювання та методів побудови систем підтримки прийняття рішень. У роботі [47] наведено приклади типових систем підтримки прийняття рішень, архітектуру систем підтримки прийняття рішень та баз даних, що використовуються у системах підтримки прийняття рішень.

Таким чином, на сьогоднішній день різними науковцями виконано дослідження стосовно розробки архітектури систем підтримки прийняття рішень для аналізу та прогнозування часових рядів, дослідження у цій галузі тривають.

1.6. Висновки до розділу 1 та постановка задач дослідження

За розділом 1 зроблено такі висновки:

1. Виконано аналіз можливостей застосування системного підходу до задач аналізу та прогнозування часових рядів. Показано, що ці задачі, а також задача розробки систем підтримки прийняття рішень для аналізу та прогнозування часових рядів є системними задачами. Під час розробки нових методів для аналізу та прогнозування часових рядів, нових систем підтримки прийняття рішень для аналізу та прогнозування часових рядів необхідно враховувати фундаментальні властивості системної методології, дотримуватись принципів формування системної методології та її практичної реалізації.

2. Виконано аналіз особливостей, які необхідно розглядати під час дослідження динаміки часових рядів. Наведено статистичні тести, що

застосовуються для аналізу ряду на нестационарність, зокрема, для виявлення тренду та гетероскедастичності. Описано подальші можливі дії дослідника у випадку виявлення нестационарності часового ряду. Виконано пояснення особливостей сезонних ефектів, що можуть впливати на часовий ряд. Наведено методи відновлення пропущених значень часових рядів.

3. Виконано аналіз найбільш поширених методів відновлення істинних закономірностей часових рядів, а саме, методу 13-місячного ковзного середнього, моделі сонячного циклу з використанням визначення параметрів сонячного циклу та різницевого рівняння.

4. Наведено огляд та виконано аналіз поширених моделей і методів, що використовуються для прогнозування часових рядів. Розглянуто регресійні методи, методи експоненційного згладжування, фільтр Калмана, методи групового урахування аргументів. Наведено переваги та недоліки цих методів, а також особливості їх застосування.

5. Виконано аналіз сучасних можливостей застосування систем підтримки прийняття рішень для аналізу та прогнозування часових рядів. Наведено детальний алгоритм побудови системи підтримки прийняття рішень та останні дослідження у цій галузі.

Постановка задач дослідження

Мета дослідження – розробка методів оцінювання параметрів математичних моделей, що змінюються за невідомим законом, та їх застосування для підвищення якості оцінок прогнозів розвитку часових рядів шляхом удосконалення системної методології прогнозування; розробка методів відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробка оригінального програмного продукту у формі системи підтримки прийняття рішень. Для досягнення мети потрібно вирішити такі завдання:

– для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створити методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два; обрати кращі методи серед створених та застосувати їх для розробки адаптивних фільтрів Калмана;

– для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створити метод оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду однозалежних псевдовимірів цього параметру; застосувати створений метод для розробки алгоритмів прогнозування на основі фільтра Калмана;

– розробити метод двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробити інтегральний критерій адекватності моделі для вибору оптимального параметра згладжування з метою застосування у методах, які пов'язані з експоненційним згладжуванням;

– розробити методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування;

– розробити метод ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробити критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів;

– розробити метод адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів;

– створити архітектуру та структурну схему системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням

часових рядів, а також виконати її програмну реалізацію мовою програмування Java.

Формалізована постановка задач дослідження

Для класу стаціонарних процесів, що мають такий вигляд:

$$\{y(k)\} \sim D(\mu, \sigma_y^2), k = 0, 1, \dots, N-1,$$

де D – довільний розподіл;

$$\mu = E[y(k)] = \text{const}; \sigma_y^2 = E[(y(k) - \mu)^2] = \text{const};$$

та класу нестационарних процесів (підкласи інтегрованих та гетероскедастичних процесів), що мають такий вигляд:

$$\{y(k)\} \sim D(\mu, \sigma_y^2), k = 0, 1, \dots, N-1,$$

$$\mu = E[y(k)] \neq \text{const}; \sigma_y^2 = E[(y(k) - \mu)^2] \neq \text{const} < \infty;$$

необхідно:

- оцінити параметри моделі;
- відновити пропущені значення часових рядів для модельних і фактичних часових рядів;
- відновити істинні закономірності для модельних і фактичних часових рядів;
- обчислити оцінки прогнозів $\hat{y}(k+p) = E_k[y(k+p)], p = 0, 1, \dots, r$ для модельних і фактичних часових рядів:

$$\{y_{\text{fact}}(k)\} = [y_1, \dots, y_N] \sim D(\bar{y}, S_y^2),$$

де p – кількість кроків прогнозування; S_y^2 – вибіркова дисперсія; \bar{y} – вибіркоче середнє.

РОЗДІЛ 2

РОЗРОБКА МЕТОДІВ ОЦІНЮВАННЯ НЕВІДОМИХ ПАРАМЕТРІВ МАТЕМАТИЧНОЇ МОДЕЛІ ТА ПОБУДОВА АДАПТИВНИХ ФІЛЬТРІВ КАЛМАНА З ВИКОРИСТАННЯМ ОТРИМАНИХ ОЦІНОК

Сучасні стохастичні процеси в економіці, фінансах, екології, технологіях і технічних системах характеризуються складною динамікою, нестационарністю (наявністю тренду і гетероскедастичності) та нелінійністю, виконання аналізу таких процесів є важливою задачею на сьогоднішній день. Тому для моделювання і прогнозування таких процесів необхідно розробляти нові та удосконалені методи, які забезпечать обчислення високоякісних прогнозів.

У процесі аналізу та прогнозування часових рядів зазвичай обробляються тільки безпосередні значення часових рядів. Проте при розв'язуванні прикладних задач велику інформаційну цінність може мати швидкість зміни значень часового ряду. Тому актуальним є використання математичної моделі у просторі станів, у якій в якості компонент вектора стану використовуються значення часового ряду та швидкість зміни його значень. Також для покращення результатів важливою є можливість оцінювання невідомих статистичних характеристик шумів наведеної моделі у просторі станів.

Точність оцінок значень параметрів у рівняннях, якими описується система, великою мірою визначає оптимальність методів оцінювання на основі фільтра Калмана. У багатьох практичних випадках знання чисельних значень усіх або кількох характеристик вказаних параметрів є неповним. Тому актуальним завданням на сьогоднішній день є знаходження оцінок невідомих характеристик параметрів системи.

2.1. Застосування системного підходу до аналізу часових рядів та розробки методів їх прогнозування

Використання системного підходу є важливим під час аналізу та прогнозування часових рядів, оскільки це дає можливість виконати декомпозицію задачі аналізу на менші задачі стільки разів, скільки це необхідно, комплексно підійти до питання застосування та розробки нових методів прогнозування часових рядів.

Під час аналізу та прогнозування часових рядів необхідно визначити параметри системи, які необхідно ідентифікувати, якщо такі параметри є та у цьому є необхідність. Ідентифікація невідомих параметрів системи дасть змогу отримати точніші оцінки значень часового ряду під час виконання прогнозування. Також необхідно визначити критерії, за якими можна буде зробити висновок про адекватність побудованих моделей та визначити кращі моделі серед тих, для яких виконувалось прогнозування.

Під час розробки нових методів прогнозування необхідно дотримуватись основних принципів формування системної методології та її практичної реалізації [12]: системної погодженості, процедурної повноти, функціональної ортогональності, інформаційної взаємозалежності, цілеспрямованої відповідності, функціональної раціональності, багатоцільової загальності, багатфакторної адаптивності, процедурної відкритості, раціональної доповнюваності. Нові створені методи аналізу та прогнозування часових рядів повинні бути універсальними, повними та становити єдину системну методологію. Структурну схему системного підходу до аналізу та прогнозування часових рядів наведено на рис. 2.1.

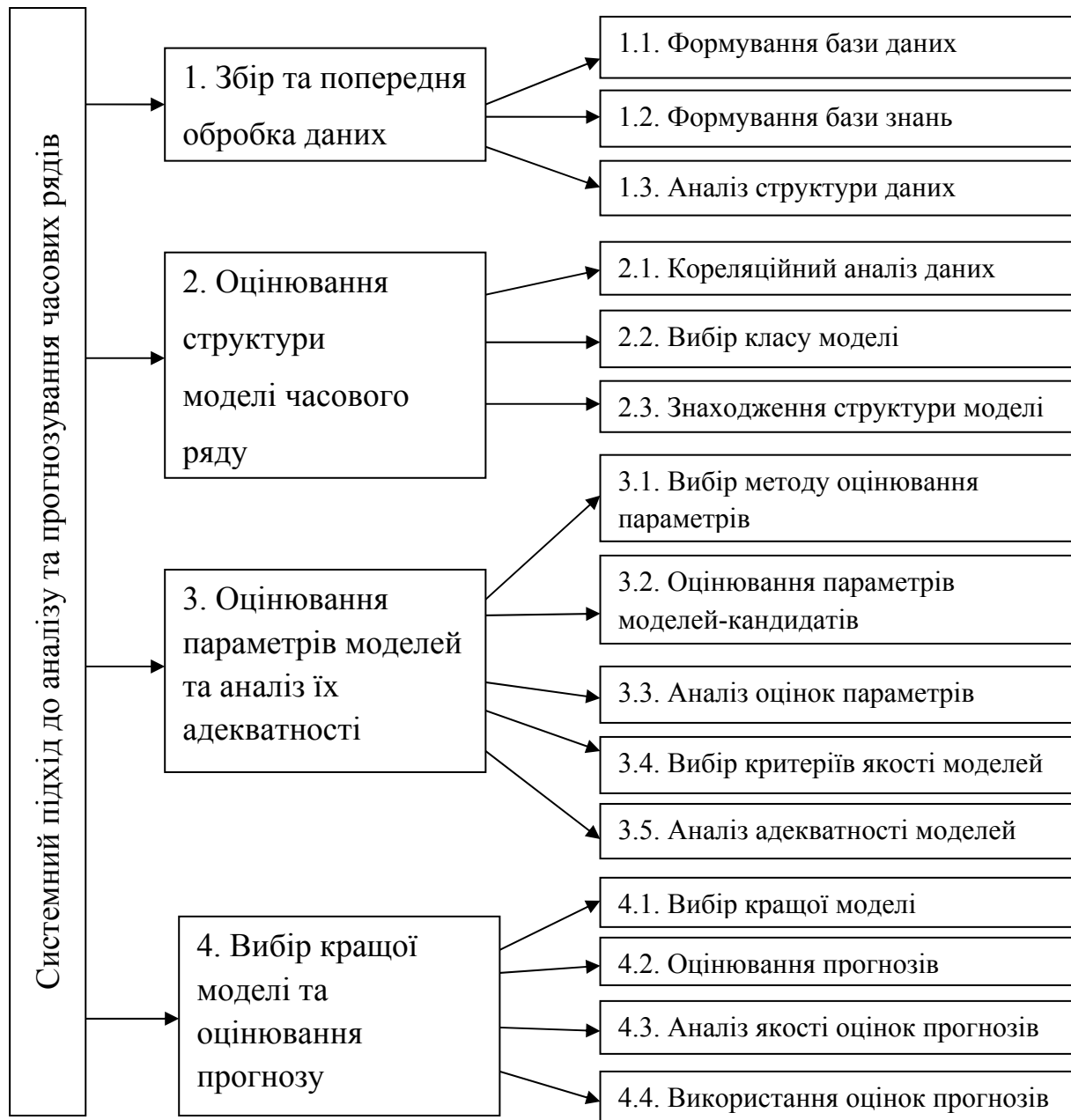


Рис. 2.1. Структурна схема системного підходу до аналізу та прогнозування часових рядів

На рис. 2.1 показано основні етапи застосування системного підходу до аналізу та прогнозування часових рядів. На першому етапі відбувається збір та попередня обробка даних, на другому етапі – оцінювання структури побудованих моделей, на третьому етапі – оцінювання параметрів моделей та аналіз їх адекватності, на четвертому етапі – аналіз отриманих результатів, вибір кращої моделі та подальше застосування оцінок прогнозу. Наведена

структурна схема описує повний комплексний підхід до аналізу та прогнозування часових рядів.

Таким чином, використання системного підходу є важливим під час аналізу та прогнозування часових рядів, оскільки це дає можливість виконати правильну декомпозицію задачі системного аналізу та виконати визначену послідовність кроків для досягнення необхідного результату.

2.2. Аналіз методів оцінювання параметрів математичної моделі та побудова адаптивних фільтрів Калмана з використанням обчислених оцінок

У процесі аналізу та прогнозування часових рядів зазвичай обробляються тільки значення часових рядів. Проте при розв'язанні прикладних задач велику інформаційну цінність може мати швидкість зміни значень часового ряду [48]. Тому у даній постановці задачі розглядається вектор стану, компонентами якого є значення часового ряду та швидкість зміни цих значень [1].

Для розв'язання задач прогнозування часових рядів використовується модель у просторі станів, статистичні характеристики шумів якої невідомі і оцінюються за реальними даними. У роботі розглядаються такі параметри цієї моделі [1]: математичне сподівання прискорення зміни значень вибірки даних, дисперсія прискорення зміни значень вибірки даних, дисперсія шуму вимірів.

Перші вагомі дослідження, що стосуються адаптивного фільтра Калмана, було зроблено у роботі [49]. Інші вчені продовжили дослідження, що стосуються адаптивного фільтра Калмана для ідентифікації збурення стану та шумів вимірів [50]. Виконувались дослідження, в яких порівнювались звичайний та адаптивний фільтри Калмана, ці дослідження показали переваги застосування адаптивного фільтра Калмана [51]. Проте подальше виконання досліджень для визначення ефективних методів оцінювання параметрів моделі для їх використання під час побудови адаптивного фільтра Калмана є

актуальним на сьогоднішній день, оскільки це може дати змогу отримувати точніші результати при прогнозуванні часових рядів.

Для оцінювання невідомих параметрів у даній роботі використовується підхід, запропонований у роботі [34]. Розглядаються нев'язки субоптимального фільтра Калмана, що будуються для моделі вільної динамічної системи з фіксованою пам'яттю. Функціями невідомих параметрів моделі є статистичні характеристики нев'язок такого фільтра. Отримані оцінки застосовуються для побудови адаптивного фільтра Калмана і розв'язання задач прогнозування.

Однією з важливих задач є прогнозування цін на кольорові метали. Прогнозування цін на цинк є важливим для тих трейдерів, що займаються купівлею та продажем металів на біржі металів. Тому аналіз та прогнозування середньодобових цін на цей метал є актуальним.

2.2.1. Математична постановка задачі

Система рівнянь, що складається з рівняння стану та рівняння вимірів, описує математичну модель, яка розглядається у даній роботі.

Рівняння стану має вигляд [34]:

$$\mathbf{X}_{i+1} = \mathbf{\Phi}\mathbf{X}_i + \mathbf{G}a_i, \quad (2.1)$$

де $\mathbf{X} = \begin{bmatrix} x_i \\ v_i \end{bmatrix}$ – вектор стану; $\mathbf{\Phi} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$ – перехідна матриця; $\mathbf{G} = \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix}$ – матриця

входів; x_i – i -е значення вибірки даних; v_i – швидкість зміни i -го значення вибірки даних; T – проміжок часу між i -м та $(i+1)$ -м станами; a_i – i -е значення прискорення зміни значень вибірки даних з такими статистичними властивостями:

$$E[a_i] = q^*, E[(a_i - q^*)(a_k - q^*)] = (\sigma_a^*)^2 \delta(i - k),$$

де q^* – математичне сподівання прискорення зміни значень вибірки даних; σ_a^* – дисперсія прискорення зміни значень вибірки даних; $\delta(i-k)$ – значення дельта-функції Дірака в $(i-k)$ -й момент часу.

Рівняння вимірів має вигляд [34]:

$$Z_i = \mathbf{H}\mathbf{X}_i + W_i, \quad (2.2)$$

де $\mathbf{H} = |1 \ 0|$ – матриця спостережень; W_i – шум вимірів у i -й момент часу з наступними статистичними властивостями:

$$E[W_i] = 0, \quad E[W_i W_k] = (\sigma^*)^2 \delta(i-k),$$

де σ^* – дисперсія шуму вимірів. Априорно невідомими вважаються статистичні характеристики q^* , σ_a^* та σ^* [1].

Субоптимальна оцінка вектора стану фільтра з фіксованою пам'яттю $m=2$ обчислюється так [34]:

$$\hat{\mathbf{X}}_i^{(2)} = \begin{bmatrix} Z_i \\ \frac{Z_i - Z_{i-1}}{T} \end{bmatrix}, i = 2, 3, \dots \quad (2.3)$$

Виконується виведення значення нев'язки субоптимального фільтра з пам'яттю $m=2$. Використовується така формула [34]:

$$\mathbf{v}_i^{(2)} = Z_i - \mathbf{H}\hat{\Phi}\hat{\mathbf{X}}_{i-1}. \quad (2.4)$$

Підставляючи у формулу (2.4) значення матриць \mathbf{H} та $\hat{\Phi}$, можна отримати [1]:

$$\begin{aligned} \nu_i^{(2)} &= Z_i - (1 \ 0) \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Z_{i-1} \\ \frac{Z_{i-1} - Z_{i-2}}{T} \end{pmatrix} = \\ &= Z_i - (1 \ T) \begin{pmatrix} Z_{i-1} \\ \frac{Z_{i-1} - Z_{i-2}}{T} \end{pmatrix} = Z_i - 2Z_{i-1} + Z_{i-2}. \end{aligned} \quad (2.5)$$

Виконане у роботі [1] доведення та отриманий результат у формулі (2.5) підтверджують формулу нев'язки субоптимального фільтра за пам'яттю $m=2$, що була отриманою у [34]:

$$v_i^{(2)} = Z_i - 2Z_{i-1} + Z_{i-2}. \quad (2.6)$$

Середнє значення нев'язки обчислюється так [34]:

$$E[v_i^{(2)}] = T^2 q^*. \quad (2.7)$$

Таким чином, математичне сподівання нев'язки субоптимального фільтра з фіксованою пам'яттю $m=2$ є функцією вимірів невідомого математичного сподівання прискорення зміни значень вибірки даних.

2.2.2. Побудова та аналіз методів оцінювання математичного сподівання прискорення зміни значень вибірки даних

Вважаються апріорно невідомими параметри математичної моделі, що описується рівняннями (2.1) – (2.2) [1]. Оцінки невідомих параметрів необхідно отримати для виконання прогнозування за цією моделлю.

Важливою задачею є отримання оцінки значення q – математичного сподівання прискорення зміни значень вибірки даних, ця оцінка позначається q^* . Для цього може бути використаною оцінка значень q^* на основі вибірки нев'язок з пам'яттю $n = m$ за такою формулою [34]:

$$\hat{q}^{(k)} = \frac{1}{kT^2} \sum_{i=1}^k v_i^{(m)}.$$

Метод № 1 полягає в оцінюванні значення q^* за повною вибіркою нев'язок з пам'яттю $n = 2$ за формулою (2.8) та підстановці отриманих значень у математичну модель (2.1)–(2.2) [34]:

$$\hat{q}^{(i)} = \frac{1}{(i-2)T^2} \sum_{j=3}^i v_j^{(2)}. \quad (2.8)$$

Метод № 2 полягає в оцінюванні значення q^* за вибірковою послідовністю нев'язок з пам'яттю $n = 2$, де взято до розгляду нев'язки, номери яких утворюють послідовність $i = 3, 5, 7, \dots$, за формулою (2.9) та підстановці отриманих значень у математичну модель (2.1)–(2.2) [34]:

$$\hat{q}^{(i)} = \frac{2}{(i-1)T^2} \sum_{j=1}^{\frac{i-1}{2}} v_{2j+1}^{(2)}. \quad (2.9)$$

Метод № 3 полягає в оцінюванні значення q^* за вибірковою послідовністю нев'язок з пам'яттю $n=2$, де взято до розгляду нев'язки, номери яких утворюють послідовність $i=3,6,9,\dots$, за формулою (2.10) та підстановці значень у математичну модель (2.1)–(2.2) [34]:

$$\hat{q}^{(i)} = \frac{3}{iT^2} \sum_{j=1}^{\frac{i}{3}} v_{3j}^{(2)}. \quad (2.10)$$

До розгляду запропоновано також метод № 4, який полягає в оцінюванні значення q^* за вибірковою послідовністю нев'язок з пам'яттю $n=2$, де взято до розгляду нев'язки, номери яких утворюють послідовність $i=3,7,11,\dots$, за формулою (2.11) та у підстановці отриманих значень у математичну модель (2.1)–(2.2) [1]:

$$\hat{q}^{(i)} = \frac{4}{(i+1)T^2} \sum_{j=1}^{\frac{i+1}{4}} v_{4j-1}^{(2)}. \quad (2.11)$$

Вибіркове середнє обчислюється таким чином [26]:

$$\bar{y} = \frac{1}{N} \times \sum_{k=1}^N y(k), \quad (2.12)$$

де \bar{y} – вибіркове середнє; N – кількість значень ряду; $y(k)$ – значення ряду в k -й момент часу.

Метод № 5 полягає в оцінюванні значення q^* за допомогою застосування до послідовності нев'язок з пам'яттю $n=2$ експоненційного згладжування зі значенням початкового наближення, обчисленим за формулою (2.12), та підстановці отриманих значень у математичну модель (2.1)–(2.2) [1]:

$$Z_q^{(i)} = \frac{1}{(i-2)T^2} \sum_{j=3}^i v_j^{(2)}, \quad (2.13)$$

$$\hat{q}^{(i)} = \alpha \times Z_q^{(i)} + (1 - \alpha) \times \hat{q}^{(i-1)}, \quad (2.14)$$

де $Z_q^{(i)}$ – ряд псевдовимірів $q^{(i)}$.

Метод № 6 полягає в оцінюванні значення q^* за послідовністю нев'язок з пам'яттю $n = 2$ за допомогою експоненційного згладжування за формулами (2.13) та (2.14) зі значенням початкового наближення, рівним першому значенню послідовності нев'язок, та у підстановці отриманих значень у математичну модель (2.1)–(2.2) [1].

Для порівняння розглянутих вище методів виконано імітаційне моделювання для моделі, що описується рівняннями (2.1)–(2.2), для випадкового процесу для кількості точок $N = 100$ та незмінного періоду часу $T = 1$, де випадкові змінні a_i та W_i визначаються за такими формулами [1]:

$$a_i = randn(1,1) \times \sqrt{(\sigma_a^*)^2} + q^*,$$

$$W_i = randn(1,1) \times \sqrt{(\sigma^*)^2}.$$

Як критерій оцінювання використовується значення дисперсії у кожний момент часу для математичного сподівання прискорення зміни значень вибірки даних, обчислене таким чином [1]:

$$D[\hat{q}^{(i)}] = [\hat{q}^{(i)} - \hat{q}^{(i-1)}]^2. \quad (2.15)$$

На рис. 2.2–2.5 наведено значення дисперсії у кожний момент часу, що розраховані за формулою (2.15), обчислені за значеннями оцінок $\hat{q}^{(i)}$, отриманими за методами № 1 – 6 [1].

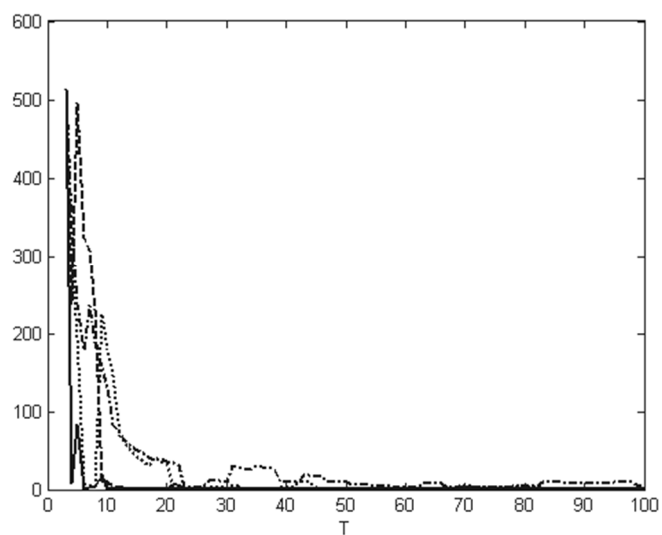


Рис. 2.2. Значення дисперсії $D[q^{(i)}]$ при початкових значеннях $q^* = 1$, $(\sigma_a^*)^2 = 10$, $(\sigma^*)^2 = 50$ для методів № 1–4, де — метод № 1, -- — метод № 2, ... — метод № 3, -.- — метод № 4

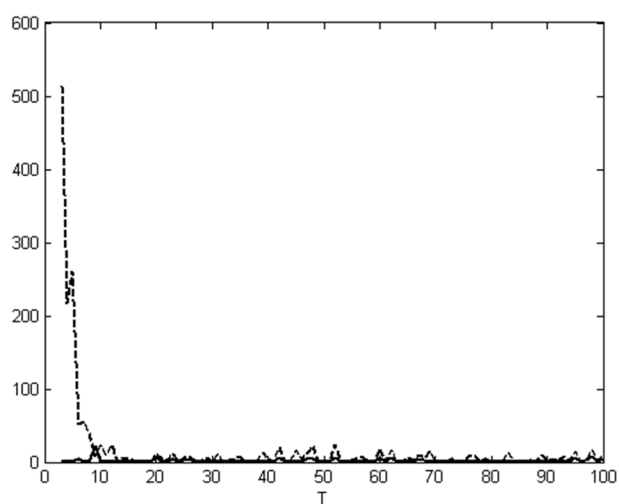


Рис. 2.3. Значення дисперсії $D[q^{(i)}]$ при початкових значеннях $q^* = 1$, $(\sigma_a^*)^2 = 10$, $(\sigma^*)^2 = 50$ для методів № 5–6, де — метод № 5, -- — метод № 6

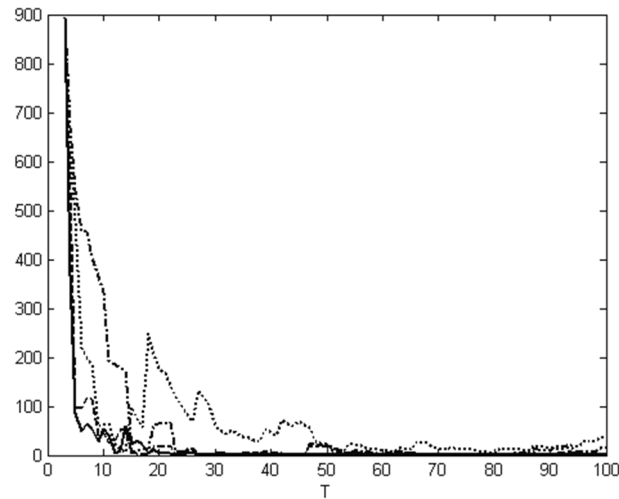


Рис. 2.4. Значення дисперсії $D[q^{(i)}]$ при початкових значеннях $q^* = 10$, $(\sigma_a^*)^2 = 50$, $(\sigma^*)^2 = 100$ для методів № 1–4, де — метод № 1, -- метод № 2, ... — метод № 3, -.- — метод № 4

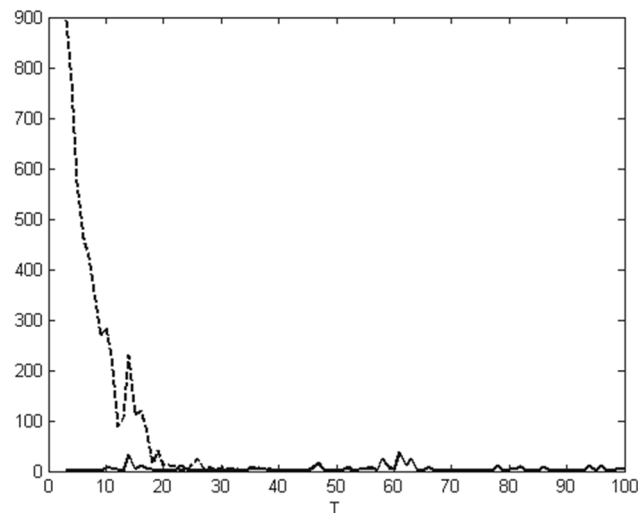


Рис. 2.5. Значення дисперсії $D[q^{(i)}]$ при початкових значеннях $q^* = 10$, $(\sigma_a^*)^2 = 50$, $(\sigma^*)^2 = 100$ для методів № 5–6, де — метод № 5, -- метод № 6

Результати виконаного імітаційного моделювання, наведені на рис. 2.2–2.5, показують, що для всіх розглянутих методів значення дисперсії

для математичного сподівання прискорення зміни значень вибірки даних збігаються до нуля за 20–40 кроків, що показує ефективність запропонованих методів. Показано, що при фіксованому значенні q^* зі збільшенням значень $(\sigma_a^*)^2$ та $(\sigma^*)^2$, а також зі збільшенням різниці між цими значеннями збіжність значень дисперсії для математичного сподівання прискорення зміни значень вибірки даних до нуля погіршується. При незмінних значеннях $(\sigma_a^*)^2$ та $(\sigma^*)^2$ зі збільшенням значення q^* збіжність значень дисперсії до нуля для математичного сподівання прискорення зміни значень вибірки даних погіршується.

Також за результатами імітаційного моделювання зроблено висновок, що кращими є методи № 1, № 5 та № 6 [1]. Показано, що оцінка значення q^* за побудованими вибірковими послідовностями нев'язок не призводить до покращення результатів порівняно з оцінкою значення q^* за повною послідовністю нев'язок [1].

Таким чином, виконано аналіз існуючих методів та створено нові методи оцінювання математичного сподівання прискорення зміни значень вибірки даних, проведено імітаційне моделювання та визначено критерій для вибору кращого методу.

2.2.3. Розробка адаптивних фільтрів Калмана та їх застосування для прогнозування середньодобових цін на цинк

З використанням отриманих у попередніх підрозділах теоретичних результатів розроблено два адаптивних фільтри Калмана.

Адаптивний фільтр Калмана № 1 [1]:

1. Спираючись на результати, отримані раніше, для побудови математичної моделі та прогнозування за нею для знаходження значень q^* використовується метод № 5.

2. Після цього знайдені значення q^* підставляються у модель, що описується рівняннями (2.1)–(2.2) та будується адаптивний фільтр Калмана, що описується такими рівняннями [1]:

екстраполяція:

$$\hat{\mathbf{X}}_{k+1} = \Phi \hat{\mathbf{X}}_k + \mathbf{B}_k, \quad (2.16)$$

$$\mathbf{P}_{k+1} = \Phi_{k+1} \mathbf{F}_k \Phi_{k+1}^T + \mathbf{G} \mathbf{Q}_k \mathbf{G}^T; \quad (2.17)$$

фільтрація:

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_{k-1} + \mathbf{K}_k \mathbf{v}_k, \quad (2.18)$$

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (2.19)$$

$$\mathbf{F}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k, \quad (2.20)$$

де $\mathbf{v}_k = Z_k - \mathbf{H}_k \hat{\mathbf{X}}_{k-1}$ – нев’язка фільтра; $\mathbf{B}_k = \begin{bmatrix} q_k \\ 0 \end{bmatrix}$; $\mathbf{Q}_k = (\sigma_a^*)^2$; $\mathbf{R}_k = (\sigma^*)^2$.

В якості початкового наближення матриці \mathbf{F}_1 обрано таку матрицю [1]:

$$\mathbf{F}_1 = \begin{bmatrix} (\sigma^*)^2 & \frac{(\sigma^*)^2}{T} \\ \frac{(\sigma^*)^2}{T} & \frac{2 \times (\sigma^*)^2}{T^2} \end{bmatrix}.$$

Критерієм оптимальності фільтра Калмана є мінімізація математичного сподівання суми квадратів похибок оцінки вектора стану [34]:

$$\min_k E\{[\mathbf{X}_k - \hat{\mathbf{X}}_k]^T [\mathbf{X}_k - \hat{\mathbf{X}}_k]\}. \quad (2.21)$$

3. При прогнозуванні часового ряду в якості оцінки значення q^* у математичну модель, що описується рівняннями (2.1)–(2.2), буде підставлено значення останньої розрахованої за відомими значеннями ряду оцінки $q^{\wedge(N)}$ [1].

Адаптивний фільтр Калмана № 2

Адаптивний фільтр Калмана № 2 є подібним до адаптивного фільтра Калмана № 1, запропонованого у роботі [1], проте на першому кроці запропоновано виконувати оцінювання прискорення зміни значень вибірки

даних, що змінюється за невідомим законом. Запропоновано значення математичного сподівання прискорення зміни значень вибірки даних $q^{(i)}$ обчислювати за методом подвійного експоненційного згладжування [9]:

$$\begin{aligned} q^{(i)} &= \alpha \times Z_q^{(i)} + (1 - \alpha) \times (q^{(i-1)} + b^{(i-1)}), i = 3, \dots, n, \\ b^{(i)} &= \beta \times (q^{(i)} - q^{(i-1)}) + (1 - \beta) \times b^{(i-1)}, i = 3, \dots, n, \end{aligned}$$

де α і β – параметри згладжування $0 < \alpha < 1$, $0 < \beta < 1$.

Отримані оцінки $q^{(i)}$ підставляються у модель та використовуються при побудові адаптивного фільтра Калмана, описаного у роботі [1], на другому кроці.

Таким чином, розроблено адаптивні фільтри Калмана № 1 та № 2, що є подібними, але відрізняються між собою методами оцінювання значення математичного сподівання прискорення зміни значень вибірки даних.

На основі вибірки даних (43 значення) про середньодобові ціни Лондонської біржі металів на цинк в американських доларах за тонну з 20.04.2017 по 21.06.2017 у робочі дні [52] необхідно виконати прогнозування на наступні п'ять робочих днів.

Виконується побудова моделей та прогнозування за допомогою адаптивних фільтрів Калмана № 1 та № 2. Також виконується побудова та прогнозування значень середньодобових цін на цинк за моделлю експоненційного згладжування, авторегресійною моделлю, авторегресійною моделлю з ковзним середнім та багаторядним поліноміальним методом групового урахування аргументів. Модель експоненційного згладжування описується формулою (1.13), початкове наближення обчислюється за формулою (2.12). Авторегресійна модель описується формулою (1.14). Авторегресійна модель з ковзним середнім описується формулою (1.20). Багаторядний поліноміальний метод групового урахування аргументів, повний опис якого наведено у розділі 1.4.4, описується формулами (1.66) – (1.69).

Після побудови математичних моделей обчислено значення прогнозних характеристик, а саме, значення середньоквадратичної похибки RMSE, середньої відсоткової абсолютної похибки MAPE та коефіцієнта Тейла U за формулами (2.22), (2.23), (2.24) відповідно. Критерієм покращення придатності побудованої моделі для прогнозування є мінімізація перерахованих вище прогнозних характеристик.

Значення середньоквадратичної похибки RMSE обчислюється так [26]:

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n [y(i) - \hat{y}(i)]^2}, \quad (2.22)$$

де n – загальна кількість днів, на які здійснювався прогноз; $y(i)$ – реальні значення; $\hat{y}(i)$ – значення, що отримані за математичними моделями; $i = 1, \dots, n$ – дні, на які здійснювався прогноз.

Значення середньої відсоткової абсолютної похибки MAPE обчислюється таким чином [26]:

$$MAPE = \frac{1}{n} \times \sum_{i=1}^n \frac{|y(i) - \hat{y}(i)|}{|y(i)|} \times 100\%. \quad (2.23)$$

Значення коефіцієнта Тейла U обчислюється так [26]:

$$U = \frac{\sqrt{\frac{1}{n} \times \sum_{i=1}^n [y(i) - \hat{y}(i)]^2}}{\sqrt{\frac{1}{n} \times \sum_{i=1}^n [y(i)]^2} + \sqrt{\frac{1}{n} \times \sum_{i=1}^n [\hat{y}(i)]^2}}. \quad (2.24)$$

У табл. 2.1 наведено прогнозні характеристики, розраховані за побудованими моделями. Адаптивні фільтри Калмана № 1 та № 2 позначено відповідно АФК(1), АФК(2) у табл. 2.1. В якості початкових значень для цієї моделі взято $(\sigma_a^*)^2 = 0,004$, $(\sigma^*)^2 = 0,004$. Модель експоненційного згладжування позначено ЕЗ у табл. 2.1. Для цієї моделі в якості початкового наближення взято середнє значення ряду та параметр згладжування обрано за сіткою значень як значення, для якого є найменшим значення середньої відсоткової абсолютної похибки MAPE. Побудовану авторегресійну модель

порядку 2 позначено $AP(2)$ у табл. 2.1. Порядок авторегресійної моделі обрано як значення, для якого значення часткової автокореляційної функції ряду є відмінними від нуля у статистичному сенсі з використанням статистики Лjung-Бокса [23]; в якості порогу значень обрано 0,01. Побудовану авторегресійну модель другого порядку з ковзним середнім порядку 2 позначено $APKC(2,2)$ у табл. 2.1. При побудові цієї моделі порядок авторегресії обрано таким, що при ньому значення часткової автокореляційної функції ряду є відмінними від нуля у статистичному значенні з використанням статистики Лjung-Бокса [23]. Порядок ковзного середнього обрано таким, при якому значення часткової автокореляційної функції залишків моделі відмінні від нуля у статистичному значенні з використанням статистики Лjung-Бокса [23], при цьому в обох випадках в якості порогу значень обрано 0,01. Багаторядний поліноміальний метод групового урахування аргументів позначено МГУА у табл. 2.1.

Таблиця 2.1

Прогнозні характеристики моделей

	АФК(1)	АФК(2)	ЕЗ	AP(2)	APKC(2,2)	МГУА
RMSE	112,7142	111,2145	141,5929	135,6811	117,3515	115,6352
MAPE	4,1562	4,1176	5,1835	4,9715	4,3005	4,2341
U	0,0216	0,0215	0,0267	0,0256	0,0220	0,0218

За наведеними у табл. 2.1 прогнозними характеристиками можна зробити такі висновки: для АФК(2) значення RMSE менше на 1,3305%, ніж для АФК(1), менше на 21,4547%, ніж для ЕЗ, менше на 18,0324%, ніж для AP(2), менше на 5,2296%, ніж для APKC(2,2), менше на 3,8229%, ніж для МГУА; для АФК(2) значення MAPE менше на 0,9287%, ніж для АФК(1), менше на 20,5633%, ніж для ЕЗ, менше на 17,1759%, ніж для AP(2), менше на 4,2529%, ніж для APKC(2,2), менше на 2,7515%, ніж для МГУА; для АФК(2) значення U менше на 0,4629%, ніж для АФК(1), менше на 19,4757%, ніж для ЕЗ, менше на

16,0156%, ніж для $AP(2)$, менше на 2,2727%, ніж для $APKC(2,2)$, менше на 1,3761%, ніж для МГУА.

Таким чином, за наведеними у табл. 2.1 прогностичними характеристиками можна зробити висновок, що за значеннями RMSE, MAPE та U показує перевагу АФК(2) перед усіма моделями. Також можна зробити висновок, що АФК(1) показує перевагу за прогностичними характеристиками перед моделями ЕЗ, $AP(2)$, $APKC(2,2)$ та МГУА. Це можна пояснити тим, що адаптивний фільтр точніше враховує статистичні характеристики двох випадкових процесів: збурення стану та шуму вимірів.

2.3. Побудова алгоритмів прогнозування на основі фільтра Калмана з використанням методів ідентифікації змінних параметрів моделі

Можливість точно оцінити параметри рівнянь, які описують систему, є важливою для отримання оптимальних алгоритмів оцінювання на основі фільтра Калмана. У більшості практичних випадків знання чисельних значень усіх або декількох характеристик вказаних параметрів є неповним. Часто може бути заданим наближений діапазон невідомих параметрів, але при цьому якість отриманих результатів може стати невисокою.

Через невизначеність параметрів моделі системи задача оцінювання стану ускладнюється, тому що необхідно одночасно оцінювати стан та параметри з використанням даних вимірів. У роботі за редакцією Леондеса [48] і статті Острема та Ейкоффа [53] наведено методи для розв'язання задачі ідентифікації параметрів. У роботі [2] описано розробку методу, що є придатним для ідентифікації постійних та змінних за невідомим законом параметрів системи.

У роботі [2] запропоновано підхід до ідентифікації змінних статистичних параметрів моделі, основою для цього підходу є формування однозалежного ряду нев'язок для допоміжного субоптимального фільтра. Андерсоном та ін. у роботі [54] для стаціонарної системи за умови вимірюваності компонент вектора стану (одиначної матриці спостережень) було застосовано підхід для

ідентифікації постійних статистичних характеристик параметрів. Для стаціонарної системи за умов, коли не всі компоненти вектора стану вимірюються, цей підхід узагальнено у роботі М.З. Згуровського та В.М. Подладчікова [34]. Метод ідентифікації постійних статистичних характеристик шумів моделі для випадку змінних перехідної матриці та матриці спостережень розроблено Т. Подладчіковою та Р. Ван Дер Лінденом у роботі [55].

У роботі [2] для ідентифікації змінних параметрів запропоновано розширення цього підходу. Розглядається математична модель, в якій компонентами вектора стану є значення часового ряду та швидкість зміни цих значень [34], що збурюється випадковим прискоренням, математичне сподівання якого змінюється за невідомим законом. У роботі [2] метою є перехід від ідентифікації постійних до ідентифікації змінних статистичних параметрів математичної моделі. У роботі [2] розроблено метод оцінювання математичного сподівання випадкового прискорення зміни значень вибірки даних. Створений метод базується на побудові часового ряду однозалежних псевдовимірів для цього параметра моделі [2]. Виконується імітаційне моделювання для дослідження ефективності розробленого методу ідентифікації.

Алгоритм прогнозування на основі фільтра Калмана будується з використанням запропонованого методу оцінювання математичного сподівання прискорення зміни значень вибірки даних [2]. Виконано імітаційне моделювання для аналізу створеного алгоритму прогнозування. Також розроблено ще один алгоритм прогнозування, в якому додатково до попереднього алгоритму оцінюється дисперсія шуму вимірів [10]. З використанням алгоритмів прогнозування на основі фільтра Калмана та методів авторегресії, авторегресії з ковзним середнім, багаторядного поліноміального методу групового урахування аргументів виконується прогнозування цін на свинець для вибірки даних середньодобових цін Лондонської біржі металів, виконується аналіз та порівняння отриманих результатів.

2.3.1. Математична постановка задачі

У роботі [2] розглядається математична модель у просторі станів, ця модель описується системою рівнянь, що містить рівняння стану та рівняння вимірів. Зміна значень вибірки даних враховується у рівнянні стану за рахунок впливу випадкових прискорень [2]. У роботі [2] розглядаються рівняння стану (2.1) та рівняння вимірів (2.2). Апріорно невідомими вважаються статистичні характеристики q^* , σ_a^* та σ^* [2].

У роботі [2] ставиться задача розробки методу оцінювання математичного сподівання q^* прискорення зміни значень вибірки даних для випадку, коли цей параметр змінюється за невідомим законом. Виконується побудова алгоритму прогнозування на основі фільтра Калмана з використанням запропонованого методу оцінювання математичного сподівання q^* [2].

2.3.2. Метод ідентифікації математичного сподівання прискорення

З рівнянь (2.1) – (2.2) випливає, що використовуючи кожні два послідовних виміри Z_i та Z_{i-1} можна побудувати оцінку вектора стану зі скінченною дисперсією похибки цієї оцінки [34]. Використовуючи допоміжний фільтр з фіксованою пам'яттю $m = 2$, оцінку вектора стану можна визначити за формулою (2.3).

У роботі [2] показано, що нев'язка фільтра з фіксованою пам'яттю $m = 2$ розраховується за формулою (2.6), обґрунтування якої докладно показано у [1]. Середнє значення нев'язки може бути представлено формулою (2.7) у вигляді лінійної функції параметра q^* , що підлягає ідентифікації. Можна зробити висновок, що за допомогою усереднення однозалежної послідовності $\frac{v_i^{(2)}}{T^2}$ можна отримати консистентну оцінку математичного сподівання q^* [34]:

$$\hat{q}_n = \frac{1}{nT^2} \sum_{i=2}^n v_i^{(2)}.$$

У роботі [2] під час виконання дослідження запропоновано будувати ряд однозалежних псевдовимірів математичного сподівання q^* прискорення зміни значень вибірки даних, де значення ряду в i -й момент часу розраховується так [2]:

$$Z_q^{(i)} = \frac{v_i^{(2)}}{T^2}, i = 3, 4, \dots, n. \quad (2.25)$$

Для оцінювання параметра q^* , що змінюється у часі, до отриманого ряду однозалежних псевдовимірів, розрахованих за формулою (2.25), у роботі [2] запропоновано застосовувати процедуру експоненційного згладжування, яка широко використовується при оцінюванні та прогнозуванні динамічних процесів за умов відсутності повної та надійної інформації про закономірності їх розвитку [55]. Експоненційне згладжування відфільтровує стохастичну компоненту нев'язок допоміжного фільтра, спричинену шумами вимірів $W_i, i = 1, 2, \dots, n$, та враховує динаміку процесу зміни q^* за рахунок того, що останній псевдовимір використовується з більшим ваговим коефіцієнтом у той час, коли вагові коефіцієнти попередніх псевдовимірів зменшуються за експоненційним законом [2].

У роботі [2] запропоновано метод, в якому оцінюється значення q^* за допомогою застосування процедури експоненційного згладжування до значень ряду однозалежних псевдовимірів з початковим наближення, що дорівнює першому значенню ряду псевдовимірів, таким чином [2]:

$$\hat{q}^{(i)} = \alpha \times Z_q^{(i)} + (1 - \alpha) \times \hat{q}^{(i-1)}, \quad (2.26)$$

де $\hat{q}^{(i)}$ – значення експоненційного середнього в i -й момент часу; α – параметр згладжування, $0 < \alpha < 1$; $\hat{q}^{(i-1)}$ – значення експоненційного середнього в $(i - 1)$ -й момент часу; $i = 3, \dots, n$ – моменти часу, на які здійснювався прогноз.

Критерієм вибору оптимального значення параметра згладжування є мінімізація значення середньої відсоткової абсолютної похибки MAPE, що обчислюється за формулою (2.23) [2]. Вибір оптимального значення параметра згладжування для подальшого використання у розробленому методі ідентифікації математичного сподівання прискорення зміни значень вибірки даних є важливим для отримання точніших оцінок за допомогою цього методу.

2.3.3. Імітаційне моделювання розробленого методу ідентифікації

У роботі [2] виконується імітаційне моделювання для дослідження ефективності розробленого методу ідентифікації математичного сподівання q^* за умов, коли цей параметр змінюється за невідомим законом. Розглянуто наведений вище метод ідентифікації, що описується формулою (2.26). Випадковий процес моделюється за рівняннями (2.1) – (2.2), для моделювання обрано значення $n = 100$ та незмінним є період часу $T = 1$ [2].

Прискорення a_i описується як випадкова величина з дисперсією $(\sigma_a^*)^2$ і математичним сподіванням $q^{(i)}$, що змінюється за косинусоїдальним законом [2]:

$$q^{(i)} = 10 \times \cos\left(\frac{\pi}{60}i\right) + 10, i = 1, 2, \dots, n.$$

Як випадкова величина з нульовим математичним сподіванням та дисперсією $(\sigma^*)^2$ моделюється шум вимірів W_i [2].

Результати імітаційного моделювання показано на рис. 2.6, що був наведений у роботі [2].

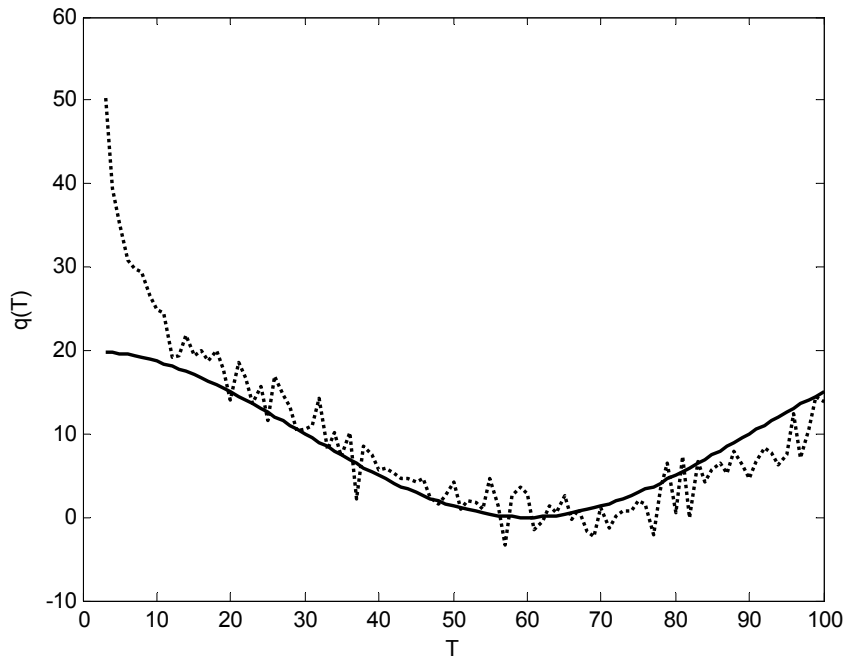


Рис. 2.6. Значення $q^{(i)}$ та $\hat{q}^{(i)}$ при істинних значеннях $(\sigma_a^*)^2 = 10$, $(\sigma^*)^2 = 50$, де
 - — істинні значення $q^{(i)}$, \cdots — оцінки $\hat{q}^{(i)}$, обчислені за розробленим методом

З рис. 2.6 видно, що зі збільшенням часу спостереження значення експоненційних середніх $\hat{q}^{(i)}$, розраховані за формулою (2.26), наближаються до істинних значень. У роботі [2] показано, що на інтервалі $T \in [3; 10]$ похибка оцінок $\hat{q}^{(i)}$ зменшується і після цього інтервалу в основному не перевищує 15% від істинних значень.

2.3.4. Алгоритми прогнозування на основі фільтра Калмана

Використання фільтра Калмана дає змогу отримати достатньо надійні оцінки за умови неповноти інформації про процес [34]. Знаходження оцінок параметрів, що подаються на вхід фільтра Калмана, надає можливість зробити кращою його роботу та отримати точніші оцінки у результаті його роботи.

Початкові дослідження стосовно оцінювання параметрів фільтра Калмана було зроблено у роботі [7]. У роботі [2] для підвищення точності прогнозу використовується розроблений метод оцінювання математичного сподівання q^* прискорення зміни значень вибірки даних, який є складовою частиною побудованих алгоритмів прогнозування на основі фільтра Калмана.

Алгоритм № 1 [2]:

1. За розробленим методом оцінювання математичного сподівання q^* виконується ідентифікація значень $q^{\wedge(i)}$ для $i = 3, 4, \dots, n$ за формулою (2.26) [2].
2. Оцінювання значення $(\sigma_a^{\wedge(i)})^2$ для $i = 4, 5, \dots, n$ за методом, наведеним у [34].

Метод оцінювання дисперсії прискорення зміни значень вибірки даних $(\sigma_a^*)^2$ містить у собі побудову стаціонарного фільтра з пам'яттю $m=2$ та з пам'яттю $m=3$ [2]. За формулою (2.6) розраховуються нев'язки фільтра з пам'яттю $m=2$. Нев'язки фільтра з пам'яттю $m=3$ розраховуються за такою формулою [34]:

$$v_i^{(3)} = Z_i - \frac{4}{3}Z_{i-1} - \frac{1}{3}Z_{i-2} + \frac{2}{3}Z_{i-3}.$$

Їх кореляційна функція обчислюється таким чином [34]:

$$C_1 = \frac{7}{12}T^4(\sigma_a^*)^2.$$

Оцінка кореляційної функції нев'язок розраховується за такою формулою [34]:

$$C_1^{\wedge} = \frac{1}{i-3} \sum_{j=4}^i (v_j^{(3)} - \frac{5}{3}q^{\wedge(j)}T^2)(v_{j-1}^{(2)} - q^{\wedge(j-1)}T^2).$$

Оцінка дисперсії прискорення зміни значень вибірки даних розраховується таким чином [34]:

$$(\sigma_a^{\wedge i})^2 = \frac{12}{7T^4}C_1^{\wedge}.$$

3. Оцінювання значення $(\hat{\sigma}^{(i)})^2$ для $i = 4, 5, \dots, n$ за підходом, наведеним у [34], з застосуванням центрованих значень нев'язок $\nu_i^{(2)}$ для оцінювання дисперсії нев'язок $\nu_i^{(2)}$ [2].

Оцінка дисперсії нев'язок $\nu_i^{(2)}$ стаціонарного фільтра при використанні прорідженої послідовності нев'язок розраховується так [2]:

$$\hat{C}_0 = \frac{2}{i-1} \sum_{j=2}^{\frac{i+1}{2}} (\nu_{2j-1}^{(2)} - q^{(2j-1)} T^2)^2.$$

Оцінка $(\hat{\sigma}^{(i)})^2$ розраховується таким чином [34]:

$$(\hat{\sigma}^{(i)})^2 = \frac{\hat{C}_0}{6} - \frac{(\hat{\sigma}_a^{(i)})^2}{12} T^4.$$

4. Побудова фільтра Калмана за формулами (2.16) – (2.21) з використанням оцінок статистичних параметрів моделі, отриманих на кроках 1–3 таким чином [2]:

$$\mathbf{B}_k = \begin{bmatrix} \hat{q}_k \\ 0 \end{bmatrix}, \quad \mathbf{Q}_k = (\hat{\sigma}_a^{(k)})^2, \quad \mathbf{R}_k = (\hat{\sigma}^{(k)})^2.$$

Початкове значення вектора $\hat{\mathbf{X}}_1$ має такий вигляд [2]:

$$\hat{\mathbf{X}}_1 = \begin{bmatrix} Z_1 \\ \frac{Z_2 - Z_1}{T} \end{bmatrix}.$$

Коваріаційна матриця похибки початкової оцінки \mathbf{F}_1 є такою [2]:

$$\mathbf{F}_1 = \begin{bmatrix} (\hat{\sigma}^{(1)})^2 & \frac{(\hat{\sigma}^{(1)})^2}{T} \\ \frac{(\hat{\sigma}^{(1)})^2}{T} & \frac{2 \times (\hat{\sigma}^{(1)})^2}{T^2} \end{bmatrix}.$$

При цьому принцип оцінювання, що є подібним до принципу оцінювання значень q^* , можна застосувати для знаходження оцінок змінних у часі значень

$(\sigma_a^*)^2$ та $(\sigma^*)^2$ у випадку, якщо розглядати в якості рядів псевдовимірів $(\hat{\sigma}_a^{(i)})^2$, $(\hat{\sigma}^{(i)})^2$ та застосувати до цих рядів процедуру експоненційного згладжування [2].

Виконується імітаційне моделювання для визначення ефективності побудованого алгоритму № 1 на основі фільтра Калмана [2]. Значення параметрів моделі, що отримуються для випадкового процесу за однією реалізацією, у цьому випадку будуть випадковими величинами і не будуть повторюватись. Для більш точного отримання результатів виконується моделювання за ансамблем реалізацій [2]. Шляхом усереднення значень, отриманих за ансамблем реалізацій [56], можна знайти статистичну дисперсію похибок оцінок значень ряду.

Для доведення ефективності алгоритму прогнозування № 1 на основі фільтра Калмана виконано імітаційне моделювання для нормально розподіленого випадкового процесу, використовується кількість точок $n = 100$, незмінний період часу $T = 1$ та $N = 100$ реалізацій [2].

Як критерій ефективності оцінювання використовується дисперсія похибок оцінок значень ряду, що описується так [2]:

$$D_i = \frac{1}{N-1} \sum_{j=1}^N [\hat{x}_j^{(i)} - x_j^{(i)}]^2, i = 3, 4, \dots, n,$$

де $x_j^{(i)}$ – істинне значення ряду в i -й момент часу при j -й реалізації; $\hat{x}_j^{(i)}$ – значення, що отримане за методом прогнозування на основі фільтра Калмана в i -й момент часу при j -й реалізації.

Результати виконаного у роботі [2] імітаційного моделювання наведено на рис. 2.7 – 2.8.

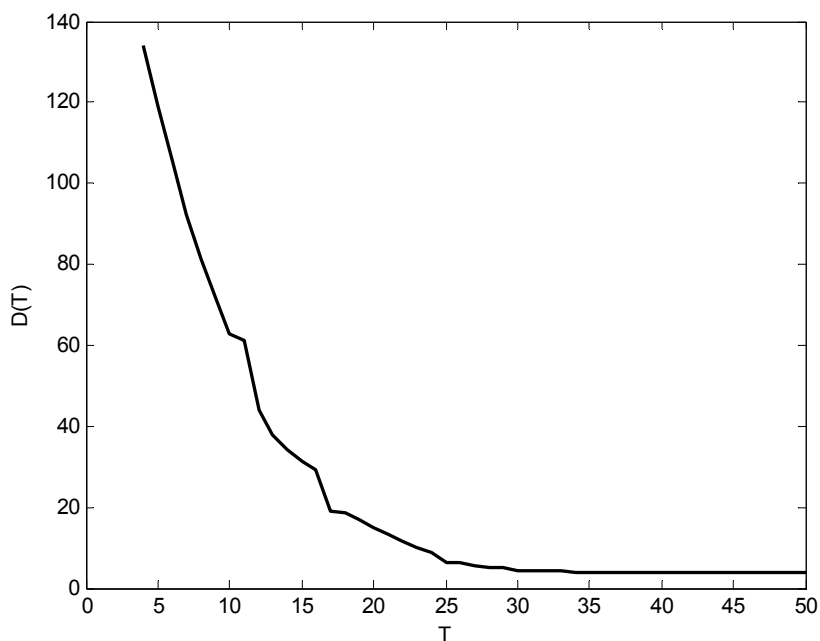


Рис. 2.7. Значення дисперсії D_i при істинних значеннях $(\sigma_a^*)^2 = 5$, $(\sigma^*)^2 = 15$

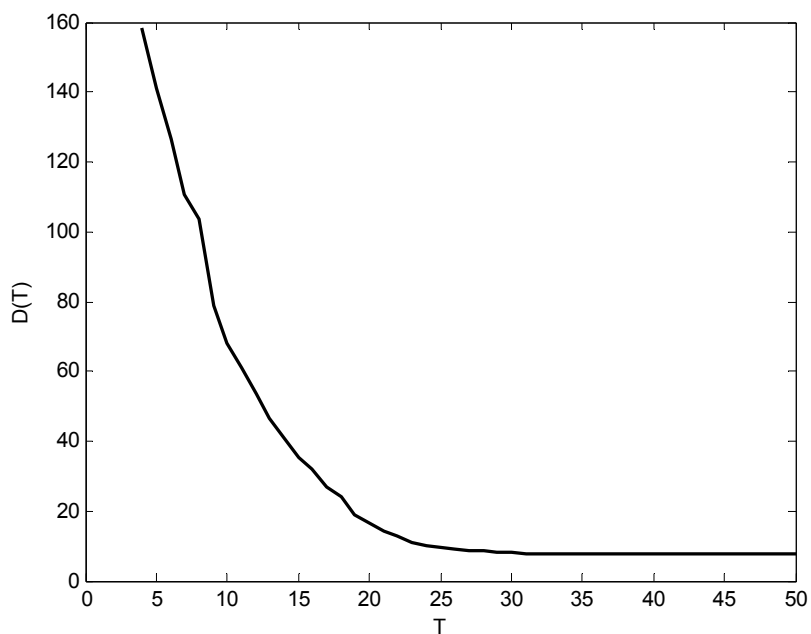


Рис. 2.8. Значення дисперсії D_i при істинних значеннях $(\sigma_a^*)^2 = 10$, $(\sigma^*)^2 = 30$

На рис. 2.7 показано, що на проміжку часу $T \in [3; 25]$ різко зменшуються значення дисперсії D_i , але вони стають практично сталими після $T = 25$ [2]. З рис. 2.8 у порівнянні з рис. 2.7 видно, що зі збільшенням істинних значень

дисперсій $(\sigma_a^*)^2$ та $(\sigma^*)^2$ зменшуються повільніше значення D_i , але після $T=28$ ці значення стають майже сталими [2]. Таким чином, наведені результати імітаційного моделювання показують для побудованого алгоритму збіжність дисперсії до сталих значень і високу точність прогнозування значень ряду.

Алгоритм № 2

Алгоритм № 2 є подібним до алгоритму № 1, описаного у роботі [2], проте на третьому кроці пропонується оцінювати значення дисперсії шуму вимірів за допомогою застосування процедури експоненційного згладжування до значень ряду незалежних псевдовимірів $(\hat{\sigma}^{(i)})^2$ зі значенням початкового наближення, рівним першому значенню ряду псевдовимірів $(\hat{\sigma}^{(i)})^2$, таким чином [10]:

$$Z_{\sigma}^{(i)} = \frac{\hat{C}_0}{6} - \frac{(\hat{\sigma}_a^{(i)})^2}{12} T^4,$$

$$(\hat{\sigma}^{(i)})^2 = \alpha \times Z_{\sigma}^{(i)} + (1 - \alpha) \times (\hat{\sigma}^{(i-1)})^2,$$

де $(\hat{\sigma}^{(i)})^2$ – значення експоненційного середнього в i -й момент часу; $Z_{\sigma}^{(i)}$ – ряд псевдовимірів $(\hat{\sigma}^{(i)})^2$; $(\hat{\sigma}^{(i-1)})^2$ – значення експоненційного середнього в $(i-1)$ -й момент часу; $i=4,5,\dots,n$ – моменти часу, на які здійснювався прогноз.

Таким чином, розроблено алгоритми прогнозування на основі фільтра Калмана № 1 та № 2, в яких використовується розроблений метод оцінювання математичного сподівання q^* прискорення зміни значень вибірки даних для підвищення точності отриманих оцінок прогнозу часових рядів. Також в алгоритмі прогнозування на основі фільтра Калмана № 2 додатково оцінюється значення дисперсії шуму вимірів, що також допомагає підвищити точність отриманих оцінок прогнозу часових рядів.

2.3.5. Прогнозування середньодобових цін на свинець

На основі вибірки даних (43 значення) про середньодобові ціни Лондонської біржі металів на свинець в американських доларах за тонну з 17.07.2017 по 14.09.2017 у робочі дні [52] виконано прогнозування на наступні п'ять робочих днів. Цей ряд даних характеризується нестационарністю, тобто він містить тренд.

Для прогнозування цін на свинець використовуються створені алгоритми прогнозування на основі фільтра Калмана, а також традиційні методи авторегресії, авторегресії з ковзним середнім, багаторядний поліноміальний метод групового урахування аргументів. Авторегресійна модель описується формулою (1.14). Авторегресійна модель з ковзним середнім описується формулою (1.20). Повний опис багаторядного поліноміального методу групового урахування аргументів наведено у розділі 1.4.4, використовуються формули (1.66) – (1.69).

Розраховуються значення прогнозних характеристик для побудованих моделей: значення середньоквадратичної похибки RMSE, середньої відсоткової абсолютної похибки MAPE та коефіцієнта Тейла U за формулами (2.22), (2.23), (2.24) відповідно. Критерієм покращення придатності побудованої моделі для прогнозування є мінімізація перерахованих вище прогнозних характеристик.

У табл. 2.2 наводяться прогнозні характеристики побудованих моделей, розраховані за формулами (2.22) – (2.24). У цій таблиці позначення виконано так: алгоритм прогнозування № 1 на основі фільтра Калмана – АПОФК(1); алгоритм прогнозування № 2 на основі фільтра Калмана – АПОФК(2); побудована за формулою (1.14) авторегресійна модель порядку 3 – АР(3); побудована за формулою (1.20) модель авторегресії порядку 3 з ковзним середнім порядку 3 – АРКС(3,3); багаторядний поліноміальний метод групового урахування аргументів – МГУА. При побудові моделей АР(3) та АРКС(3,3) порядок авторегресії обрано таким, що при цьому порядку значення часткової автокореляційної функції ряду відмінні від нуля у статистичному

сміслі з використанням статистики Лjunga-Бокса [23], поріг значень обрано як 0,01. Під час побудови моделі АРКС(3,3) порядок ковзного середнього обрано так, що при цьому порядку значення часткової автокореляційної функції залишків моделі відмінні від нуля у статистичному смислі з використанням статистики Лjunga-Бокса [23], при цьому в якості порогу значень використовується 0,01.

Таблиця 2.2

Прогнозні характеристики моделей

	АПОФК(1)	АПОФК(2)	АР(3)	АРКС(3,3)	МГУА
RMSE	114,2176	112,4524	116,9278	123,5582	115,8724
MAPE	4,3465	4,2352	4,4499	4,7844	4,3987
U	0,0245	0,0239	0,025	0,0264	0,0248

За наведеними у табл. 2.2 прогнозними характеристиками можна зробити такі висновки: для АПОФК(2) значення RMSE менше на 1,5455%, ніж для АПОФК(1), менше на 3,8275%, ніж для АР(3), менше на 8,9883%, ніж для АРКС(3,3), менше на 2,9515%, ніж для МГУА; для АПОФК(2) значення MAPE менше на 2,5607%, ніж для АПОФК(1), менше на 4,8248%, ніж для АР(3), менше на 11,4789%, ніж для АРКС(3,3), менше на 3,717%, ніж для МГУА; для АПОФК(2) значення U менше на 2,4489%, ніж для АПОФК(1), менше на 4,4%, ніж для АР(3), менше на 9,4697%, ніж для АРКС(3,3), менше на 3,629%, ніж для МГУА.

Таким чином, у табл. 2.2 показано, що значення усіх прогнозних характеристик є меншими для АПОФК(2), ніж для АПОФК(1), АР(3), АРКС(3,3) та МГУА. Це свідчить про те, що алгоритм прогнозування № 2 на основі фільтра Калмана показує перевагу перед іншими методами прогнозування. Також АПОФК(1) показує перевагу за значеннями усіх прогнозних характеристик перед АР(3), АРКС(3,3) та МГУА.

2.4. Висновки до розділу 2

Наведено важливість застосування системного підходу до аналізу та прогнозування часових рядів. Показано необхідність ідентифікації невідомих параметрів системи у випадку, якщо система їх містить. Показано необхідність дотримання основних принципів формування системної методології та її практичної реалізації під час розробки нових методів прогнозування. Створено структурну схему, на якій показано основні етапи застосування системного підходу до аналізу та прогнозування часових рядів.

У даній роботі досліджено математичну модель у просторі станів, в якій компонентами вектора стану є значення часового ряду та швидкість зміни цих значень. Використано модель у просторі станів, статистичні характеристики шумів якої невідомі і підлягають оцінюванню за фактичними даними. Розглянуто відомі методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два та запропоновано три нових методи оцінювання цього параметра. Виконано імітаційне моделювання та аналіз його результатів, за яким визначено кращі методи оцінювання. За результатами імітаційного моделювання зроблено висновок, що кращими є методи № 1, № 5 та № 6.

З використанням створених методів оцінювання математичного сподівання прискорення зміни значень вибірки даних розроблено два адаптивних фільтра Калмана. Виконано прогнозування середньодобових цін на цинк за даними Лондонської біржі металів за допомогою розроблених адаптивних фільтрів Калмана та методів експоненційного згладжування, авторегресії, авторегресії з ковзним середнім, багаторядного поліноміального методу групового урахування аргументів. За значеннями обчислених прогнозних характеристик RMSE, MAPE та коефіцієнта Тейла U розроблені адаптивні фільтри Калмана є кращими порівняно з іншими методами.

Також розроблено метод оцінювання математичного сподівання прискорення зміни значень вибірки даних для випадку, коли цей параметр змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду однозалежних псевдовимірів цього параметру. З використанням запропонованого методу оцінювання математичного сподівання прискорення зміни значень вибірки даних виконано побудову алгоритму прогнозування на основі адаптивного фільтра Калмана. Також розроблено ще один алгоритм прогнозування, в якому додатково до попереднього алгоритму виконується оцінювання дисперсії шуму вимірів. Виконано прогнозування цін на свинець з використанням двох розроблених алгоритмів прогнозування та методів авторегресії, авторегресії з ковзним середнім, багаторядного поліноміального методу групового урахування аргументів. За значеннями прогнозних характеристик RMSE, MAPE та U розроблені алгоритми прогнозування є кращими порівняно з іншими методами, що використовувались для прогнозування.

РОЗДІЛ 3

РОЗРОБКА МЕТОДІВ ВІДНОВЛЕННЯ ПРОПУЩЕНИХ ЗНАЧЕНЬ, ІСТИННИХ ЗАКОНОМІРНОСТЕЙ ТА ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ НА ОСНОВІ ДВОБІЧНОГО ЕКСПОНЕНЦІЙНОГО ЗГЛАДЖУВАННЯ

Пошук прихованих закономірностей динамічних процесів є важливою задачею на сьогоднішній день, тому що це дає змогу використовувати для подальшої роботи оцінки значень ряду, що описуються певним законом і не містять великих шумів стану та вимірів. Використання такого підходу є актуальним для специфічних сонячних та океанографічних даних. Відновлення пропущених значень часових рядів є корисним для подальшої обробки таких рядів, наприклад, для прогнозування.

Експоненційне згладжування є одним з найпоширеніших методів відновлення процесу в умовах відсутності повної інформації про закономірності розвитку процесу. Багатьма вченими розроблено різні методи на основі експоненційного згладжування, проте можливості вдосконалення та створення нових методів на цьому не вичерпано. Покращення результатів відновлення істинних закономірностей, пропущених значень та прогнозування часових рядів можна отримати шляхом використання додаткової інформації про часові ряди.

Процедура експоненційного згладжування у кожній точці є фільтрацією похибок вимірів на основі попередньої по відношенню до цієї точки інформації про виміри. Проте більше інформації про процес можна отримати, модифікуючи метод експоненційного згладжування шляхом додавання другого етапу для додаткової фільтрації похибок у всіх попередніх точках.

Додаткову інформацію про динаміку часового ряду можна отримати, аналізуючи дані про взаємозалежні часові ряди. Цю інформацію можна отримати шляхом побудови кореляційної матриці для змінних, стосовно яких відбувається дослідження.

Також додатково можна покращити результати шляхом введення ковзних інтервалів під час використання процедур експоненційного згладжування для даних, що мають специфічну періодичність, наприклад, для сонячних даних. Оцінка величини необхідного ковзного інтервалу для різних даних також є важливою задачею.

Саме тому актуальними на сьогоднішній день є подальші дослідження стосовно задач відновлення істинних закономірностей, пропущених значень та прогнозування часових рядів.

3.1. Розробка методу двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів

Пошук прихованих закономірностей та відновлення пропущених значень часових рядів є важливим етапом аналізу та прогнозування часових рядів. Задачі побудови математичних моделей, виявлення закономірностей процесів за експериментальними даними прямо або опосередковано пов'язані з їх згладжуванням та фільтрацією, що дає змогу покращити їх статистичні характеристики [23, 57].

Експоненційне згладжування є одним з найбільш поширених методів відновлення процесу в умовах відсутності повної та надійної інформації про закономірності його розвитку. Метод експоненційного згладжування був вперше запропонований Брауном [30] і згодом розширений Холтом [31]. Гарднер у своїй роботі [32] виконує аналіз теоретичних та емпіричних аргументів на користь вибору параметра згладжування, окрім цього, він розглянув різноманітні моделі експоненційного згладжування у термінах сезонності (відсутня, адитивна сезонність, мультиплікативна) і тренду (відсутній, лінійний тренд, експоненційний, демпфований). Гарднер у своїй більш пізній роботі [33] описує останні дослідження стосовно експоненційного згладжування, зокрема, він називає повне статистичне обґрунтування

експоненційного згладжування на основі нового класу моделей з одним джерелом похибок найважливішим теоретичним досягненням. У роботі [58] Гайндман та інші розглядають три класи моделей у просторі станів та застосовують ці класи для побудови аналітичних виразів для дисперсії похибки прогнозу, які можуть бути використані для побудови прогнозних інтервалів на один або декілька кроків вперед. Коласса у роботі [59] запропонував застосовувати зважені комбінації прогнозу з використанням інформаційного критерію Акайке.

У роботі [60] описано квазіоптимальну процедуру згладжування, у цій процедурі пропонується враховувати попередню інформацію про виміри під час застосування процедури експоненційного згладжування. Актуальною задачею є виконання досліджень стосовно створення нових методів відновлення закономірностей розвитку та пропущених значень часових рядів, тому що це дасть змогу виконувати поглиблену попередню обробку даних та розв'язувати завдання подальшого прогнозування часових рядів.

Також важливим є відновлення пропущених значень цін на кольорові метали для компаній, які займаються продажем та купівлею на біржі металів. Тому досить важливим практичним завданням є точне відновлення цін на метали.

3.1.1. Опис методу двобічного експоненційного згладжування

Експоненційне згладжування можна представити у вигляді лінійного фільтра, в якому послідовно на вхід надходять значення ряду, а на виході формуються значення експоненційного середнього [3]. Процедуру експоненційного згладжування у кожній точці можна представити як фільтрацію шумів вимірів за даними про попередню по відношенню до цієї точки інформацію про виміри. Проте на цьому не вичерпуються можливості більш точного відновлення ряду. Початкові дослідження стосовно створення

методу двобічного експоненційного згладжування виконано у роботі [8], далі їх продовжено і розширено у роботі [3].

У роботі [3] розглядається процедура експоненційного згладжування для часового ряду, що описується формулою (1.13), зі значенням початкового наближення, що дорівнює першому значенню ряду.

Для визначення оптимального значення параметра згладжування розглядаються такі статистичні характеристики, що використовуються для оцінювання якості побудованої моделі у цілому [3]: коефіцієнт детермінації R^2 , сума квадратів похибок моделі SSE , статистика Дарбіна-Уотсона DW .

Коефіцієнт детермінації R^2 обчислюється так [23]:

$$R^2 = \frac{\hat{Var}[y(k)]}{Var[y(k)]}, \quad (3.1)$$

де $\hat{Var}[y(k)]$ – дисперсія основної змінної $y(k)$, оціненої за моделлю; $Var[y(k)]$ – фактична дисперсія основної змінної $y(k)$, обчислена на основі вибірки даних.

Сума квадратів похибок моделі SSE обчислюється таким чином [23]:

$$SSE = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2, \quad (3.2)$$

де N – кількість значень у вибірці даних.

Статистика Дарбіна-Уотсона DW обчислюється за такою формулою [23]:

$$DW = 2 - 2 \times \rho, \quad (3.3)$$

де ρ – коефіцієнт автокореляції для $e(k)$ та $e(k-1)$, тобто при зміщенні $S=1$.

У формулі (3.3) коефіцієнт автокореляції ρ для $e(k)$ і $e(k-1)$ обчислюється таким чином [23]:

$$\rho = \frac{\sum_{k=2}^N e(k) \times e(k-1)}{\sum_{k=1}^N [e(k)]^2},$$

де $e(k)$ – похибка моделі у k -й момент часу; $e(k-1)$ – похибка моделі у $(k-1)$ -й момент часу.

У роботі [3] розроблено інтегральний критерій адекватності моделі, який описується так [3]:

$$I = |1 - R^2| + |2 - DW|,$$

де R^2 – коефіцієнт детермінації, що обчислюється за формулою (3.1); DW – статистика Дарбіна-Уотсона, що обчислюється за формулою (3.3).

У даній роботі виконано подальші дослідження та запропоновано також вдосконалення наведеного вище інтегрального критерію адекватності моделі, що описується таким чином:

$$I = \gamma \times |1 - R^2| + \eta \times |2 - DW|, \quad (3.4)$$

де R^2 – коефіцієнт детермінації, що обчислюється за формулою (3.1); DW – статистика Дарбіна-Уотсона, що обчислюється за формулою (3.3); γ, η – вагові коефіцієнти. У даній роботі вагові коефіцієнти мають такі значення: $\gamma = 1; \eta = 0,5$.

В якості критерію вибору оптимального значення параметра згладжування запропоновано у роботі [3] використовувати мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються, відповідно, за формулами (3.2) та (3.4).

У кожній точці процедура експоненційного згладжування є фільтрацією похибок вимірювання на основі попередньої по відношенню до цієї точки інформації про виміри [3]. Більш точну інформацію про процес можна отримати за рахунок модифікації методу експоненційного згладжування шляхом додавання другого етапу для додаткової фільтрації похибок у всіх попередніх точках [3].

У роботі [3] до експоненційних середніх, отриманих за формулою (1.13), запропоновано застосовувати експоненційне згладжування у зворотному часі $k = N - 1, \dots, 1$. Кожне уточнене значення S_k^* обчислюється так [3]:

$$S_k^* = \alpha \times S_k + (1 - \alpha) \times S_{k+1}^*, \quad (3.5)$$

де S_k^* – значення експоненційного середнього у зворотному часі у k -й момент часу; α – параметр згладжування; S_k – значення експоненційного середнього у k -й момент часу; S_{k+1}^* – значення експоненційного середнього у зворотному часі у $(k+1)$ -й момент часу, тут значення $k = N-1, N-2, \dots, 1$, початкове значення $S_N^* = S_N$.

У роботі [3] під час виконання згладжування у зворотному часі запропоновано в якості критерію вибору оптимального значення параметра згладжування використовувати мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються за формулами (3.2) та (3.4), відповідно, але при цьому в якості вхідних значень при обчисленні цих характеристик використовуються значення S_k . Під час розрахунку уточненої оцінки S_k^* ваги оцінок S_k надходять на вхід алгоритму у зворотному часі [3].

Таким чином, розроблено метод двобічного експоненційного згладжування, інтегральний критерій адекватності моделі та критерій вибору оптимального значення параметра згладжування. Додавання другого етапу для додаткової фільтрації похибок у всіх попередніх точках у методі двобічного експоненційного згладжування дозволяє отримати точніші оцінки значень часового ряду.

3.1.2. Імітаційне моделювання для дослідження ефективності методу двобічного експоненційного згладжування

Виконується імітаційне моделювання розробленого методу двобічного експоненційного згладжування і порівняння з методом експоненційного згладжування для дослідження ефективності розробленого методу.

Розглядається процес випадкового блукання, що описується такими рівняннями у просторі станів. Рівняння стану описується так [34]:

$$x(i+1) = x(i) + v(i), \quad (3.6)$$

де $x(i)$ – стан в i -й момент часу; $v(i)$ – шум стану в i -й момент часу з такими статистичними характеристиками:

$$E[v(i)] = q^*, E[(v(i) - q^*)(v(k) - q^*)] = \sigma^2 \delta(i - k),$$

де σ^2 – дисперсія шуму стану; $\delta(i - k)$ – значення дельта-функції Дірака в $(i - k)$ -й момент часу.

За умов наявності адитивного шуму вимірюються значення вибірки даних, рівняння вимірів має такий вигляд [34]:

$$y(i) = x(i) + w(i), \quad (3.7)$$

де $y(i)$ – вимір в i -й момент часу; $w(i)$ – шум вимірів в i -й момент часу з такими статистичними характеристиками:

$$E[w(i)] = 0, E[w(i) \times w(k)] = (\sigma^*)^2 \delta(i - k),$$

де $(\sigma^*)^2$ – дисперсія шуму вимірів. Вважаються апіорно невідомими статистичні характеристики σ та σ^* [3].

Імітаційне моделювання виконується для моделі, яка описується рівняннями (3.6) – (3.7), кількість точок для моделювання обрано $N = 50$ [3]. Шум стану $v(i)$ моделюється як випадкова величина з математичним сподіванням q^* та дисперсією σ^2 [3]. Шум виміру $w(i)$ моделюється як випадкова величина з нульовим математичним сподіванням та дисперсією $(\sigma^*)^2$ [3].

Результати імітаційного моделювання показано на рис. 3.1 [3]. Для оцінювання якості побудованих моделей у цілому обчислюються такі статистичні характеристики: коефіцієнт детермінації R^2 за формулою (3.1); сума квадратів похибок моделі SSE за формулою (3.2); статистика Дарбіна-Уотсона DW за формулою (3.3).

У табл. 3.1 наведено обчислені статистичні характеристики, які отримані за методами експоненційного та двобічного експоненційного згладжування [3]. У табл. 3.1 двобічне експоненційне згладжування позначено як ДЕЗ, експоненційне згладжування – ЕЗ.

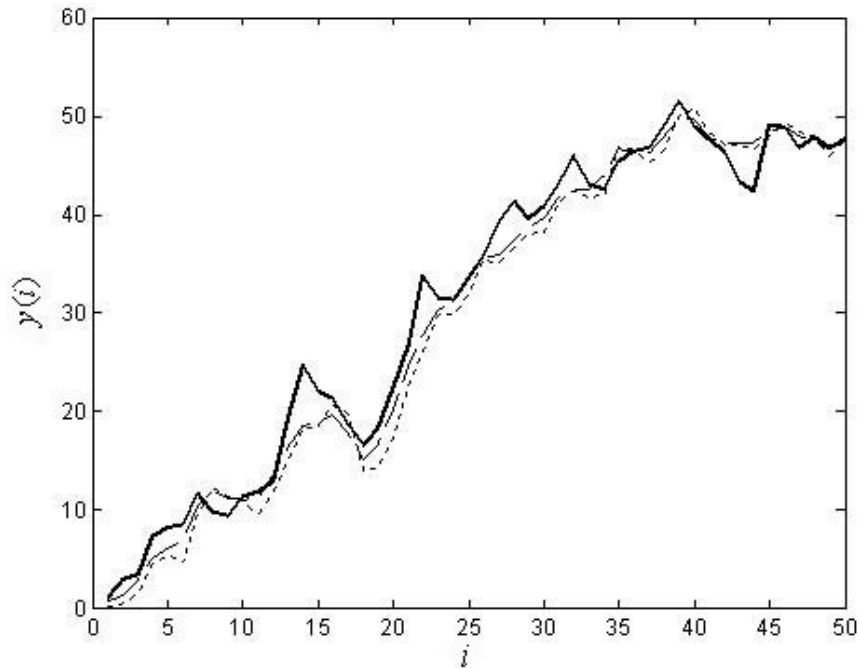


Рис. 3.1. Значення $y(i)$ та $\hat{y}(i)$ при значеннях $q^* = 1$, $\sigma^2 = 8$, $(\sigma^*)^2 = 16$, $x(1) = 1$, де символ $-$ – істинні значення $y(i)$; \cdots – оцінки $\hat{y}(i)$, отримані за методом експоненційного згладжування; $- -$ – оцінки $\hat{y}(i)$, отримані за методом двобічного експоненційного згладжування

Таблиця 3.1

Статистичні характеристики моделей

Статистична характеристика	ДЕЗ	ЕЗ
R^2	0,997	0,993
SSE	228,463	393,327
DW	0,844	0,626

З рис. 3.1 і табл. 3.1 випливає, що двобічне експоненційне згладжування дозволяє отримати оцінки значень ряду $\hat{y}(i)$, що є ближчими до вимірів та менш варіабельними, за всіма статистичними характеристиками показує

перевагу перед експоненційним згладжуванням [3]. За наведеними у табл. 3.1 статистичними характеристиками можна зробити такі висновки: коефіцієнт детермінації R^2 для двобічного експоненційного згладжування є більшим на 0,401% та ближчим до ідеального значення 1, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для двобічного експоненційного згладжування є меншою на 41,915%, ніж для експоненційного згладжування; для методу двобічного експоненційного згладжування статистика Дарбіна-Уотсона DW є більшою на 25,829% та ближчою до ідеального значення 2, ніж для методу експоненційного згладжування.

3.1.3. Застосування методу двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів

У даній роботі метод двобічного експоненційного згладжування запропоновано застосовувати для прогнозування значень часових рядів таким чином: необхідно використовувати процедуру, що описується рівняннями (1.13) та (3.5), при цьому значення параметра згладжування для процедури згладжування у прямому часі використовується таке ж, як і для всіх відомих значень, в якості початкового значення використовується останнє відоме значення.

Метод двобічного експоненційного згладжування запропоновано застосовувати для відновлення істинних закономірностей часових рядів шляхом застосування цього методу до значень на часовому інтервалі, на якому задані ці значення, відповідно до процедури, що описується рівняннями (1.13) та (3.5). Для знаходження параметра згладжування на кожному з етапів двобічного експоненційного згладжування застосовується критерій мінімізації значень, що описуються рівняннями (3.2) та (3.4).

Важливим на сьогоднішній день є відновлення пропущених значень часових рядів, яке виконується на етапі аналізу та обробки даних. Можливість

коректно виконувати подальшу обробку ряду, а саме, прогнозування, дає отримання точних відновлених значень ряду замість пропущених. У роботі [3] запропоновано алгоритм для відновлення пропущених значень часового ряду з використанням методу двобічного експоненційного згладжування.

Алгоритм № 1 описується так [3]:

1. Визначається параметр згладжування. У випадку, коли перед пропуском знаходиться від 1 до 3 значень ряду, для знаходження параметра згладжування на кожному з етапів двобічного експоненційного згладжування за критерієм мінімізації значень, що описуються рівняннями (3.2) та (3.4), запропоновано брати значення ряду, що знаходяться після пропущеного значення. У випадку, коли перед пропуском знаходиться більше трьох значень ряду, для знаходження параметра згладжування на кожному з етапів двобічного експоненційного згладжування за критерієм мінімізації значень, що описуються рівняннями (3.2) та (3.4), запропоновано брати значення ряду, що знаходяться перед пропущеним значенням.

2. Зі значенням початкового наближення, рівним середньому арифметичному серед усіх відомих значень, та з використанням в якості відомого значення виміру останнього значення перед пропущеним значенням запропоновано застосовувати процедуру двобічного експоненційного згладжування, що описується рівняннями (1.13) та (3.5) відповідно.

Таким чином, метод двобічного експоненційного згладжування може бути ефективно застосованим для прогнозування та відновлення істинних закономірностей часових рядів, а також відновлення пропущених значень часових рядів за наявності різних початкових умов.

3.1.4. Відновлення пропущених значень цін на цинк

У роботі [3] на основі вибірки даних (43 значення) щодо середньодобових цін Лондонської біржі металів на цинк в американських доларах за тону з

01.06.2016 по 31.07.2016 у робочі дні [52] виконується дослідження стосовно штучного створення пропусків та відновлення пропущених значень.

Пропущені значення цін на цинк відновлюються за допомогою таких методів [3]: розробленого методу двобічного експоненційного згладжування, що описується формулами (1.13) та (3.5); заміни пропущених значень середніми арифметичними значеннями за формулою (1.12); відновлення за допомогою методу експоненційного згладжування за формулою (1.13) з використанням в якості відомого значення виміру останнього значення перед пропущеним значенням та зі значенням початкового наближення, рівним середньому арифметичному [23] серед усіх відомих значень; за допомогою підходу, запропонованого у роботі [24], при якому прогнозування та відновлення на s кроків вперед можна виконати за такою формулою [24]:

$$\hat{y}(k+s) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k), \quad (3.8)$$

де a_0 , a_1 – коефіцієнти авторегресійної моделі першого порядку; $\hat{y}(k+s)$ – оцінка значення у $(k+s)$ -й момент часу.

Розглядаються два випадки, що відповідають ситуаціям, описаним у алгоритмі № 1 для двобічного експоненційного згладжування, та з використанням описаних вище чотирьох методів відновлення для цих випадків виконується відновлення пропущених значень часового ряду [3]. Якість побудованих моделей у цілому оцінюється за допомогою таких статистичних характеристик: коефіцієнта детермінації R^2 за формулою (3.1); суми квадратів похибок моделі SSE за формулою (3.2); статистики Дарбіна-Уотсона DW за формулою (3.3).

Випадок № 1

У вибірці даних щодо середньодобових цін на цинк зроблено штучний пропуск 5 значень з номерами $N = 2, 3, 4, 5, 6$ та виконується відновлення за чотирма описаними вище методами [3].

У табл. 3.2 наведено статистичні характеристики, що показують якість побудованих моделей у цілому, розраховані за формулами (3.1), (3.2), (3.3) відповідно на проміжку $N=1,...,43$ [3]. У цій таблиці моделі позначено так: заміна пропущених значень середніми арифметичними значеннями – СА; авторегресійний підхід, описаний у роботі [24] – АРП; експоненційне згладжування – ЕЗ та двобічне експоненційне згладжування – ДЕЗ.

Таблиця 3.2

Статистичні характеристики моделей

Статистична характеристика	СА	АРП	ЕЗ	ДЕЗ
R^2	0,868	0,768	0,828	0,888
SSE	7081,938	8218,398	7163,373	7061,476
DW	0,819	0,649	0,655	0,829

За наведеними у табл. 3.2 статистичними характеристиками можна зробити такі висновки: коефіцієнт детермінації R^2 є ближчим до ідеального значення 1 та більшим для двобічного експоненційного згладжування на 2,252%, ніж для середнього арифметичного, більшим на 13,514%, ніж для авторегресійного підходу, більшим на 6,757%, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для двобічного експоненційного згладжування є меншою на 0,289%, ніж для середнього арифметичного, є меншою на 14,077%, ніж для авторегресійного підходу, є меншою на 1,422%, ніж для експоненційного згладжування; статистика Дарбіна-Уотсона DW для двобічного експоненційного згладжування є ближчою до ідеального значення 2 та більшою на 1,206%, ніж для середнього арифметичного, більшою на 21,713%, ніж для авторегресійного підходу, більшою на 20,989%, ніж для експоненційного згладжування.

Випадок № 2

У вибірці даних щодо середньодобових цін на цинк зроблено штучний пропуск 5 значень з номерами $N = 30, 31, 32, 33, 34$ та виконується відновлення за чотирма описаними вище методами [3]. У табл. 3.3 наведено статистичні характеристики моделей аналогічно до табл. 3.2 [3].

Таблиця 3.3

Статистичні характеристики моделей

Статистична характеристика	СА	АРП	ЕЗ	ДЕЗ
R^2	0,925	0,865	0,914	0,937
SSE	1297,453	2065,189	1453,657	1132,567
DW	0,787	0,385	0,768	0,797

За наведеними у табл. 3.3 статистичними характеристиками можна зробити такі висновки: коефіцієнт детермінації R^2 є ближчим до ідеального значення 1 та більшим для двобічного експоненційного згладжування на 1,281%, ніж для середнього арифметичного, більшим на 7,684%, ніж для авторегресійного підходу, більшим на 2,455%, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для двобічного експоненційного згладжування є меншою на 12,708%, ніж для середнього арифметичного, є меншою на 45,159%, ніж для авторегресійного підходу, є меншою на 22,088%, ніж для експоненційного згладжування; статистика Дарбіна-Уотсона DW для двобічного експоненційного згладжування є ближчою до ідеального значення 2 та більшою на 1,255%, ніж для середнього арифметичного, більшою на 51,694%, ніж для авторегресійного підходу, більшою на 3,639%, ніж для експоненційного згладжування.

Розроблений метод двобічного експоненційного згладжування показує перевагу перед усіма іншими методами за значеннями статистичних

характеристик, наведених у табл. 3.2 і табл. 3.3. Таким чином, для відновлення пропущених значень часових рядів для різних початкових умов показано ефективне використання методу двобічного експоненційного згладжування.

3.2. Розробка методів відновлення пропущених значень та прогнозування для взаємозалежних часових рядів

Для багатьох напрямів досліджень задачі моделювання та прогнозування взаємозалежних процесів за умови наявності пропусків даних є актуальними. Одним з таких прикладів є дослідження стану різних країн світу з точки зору сталого розвитку. Показники сталого розвитку країн світу дають можливість оцінювати стан країн у різних аспектах, наприклад, стан економіки, медицини, екології. Тому дослідження показників сталого розвитку є важливим на сьогоднішній день. Також важливим є виконання досліджень стосовно взаємозалежних показників сталого розвитку. Це дає змогу отримати точнішу інформацію про стан певної сфери діяльності країни, ніж дослідження стосовно окремих показників.

Макейтайте та інші у своїй роботі [61] визначили шість напрямів сталого розвитку: соціальний, екологічний, економічний, культурний, архітектурний та технічний. У роботі [61] показано, що кожен з цих напрямів складається з багатьох компонентів. Межа між охороною здоров'я та соціальним забезпеченням стає все менш чіткою, що показується у роботах [62] і [63]. Індекс кількості лікарняних ліжок є одним з важливих показників сталого розвитку. Як індикатор надання послуг у сфері охорони здоров'я протягом тривалого часу розглядалась кількість лікарняних ліжок. Крістенсен та інші у роботі [64] показали, що великі лікарні (більше, ніж 400–600 лікарняних ліжок) є менш ефективними з точки зору витрат і не обов'язково є кращими у наданні медичних послуг, ніж менші лікарні. Збільшення кількості лікарняних ліжок є ефективним до того моменту, поки це призводить до покращення якості медичних послуг і при цьому допустимо збільшуються витрати на медицину.

Автори показали у роботі [64] на прикладі розрахунків для Данії, що оптимальним можна вважати функціонування лікарень, в яких менше, ніж 230 лікарняних ліжок. Збільшення кількості лікарняних ліжок також пов'язане зі збільшенням кількості хронічних захворювань та старінням населення Землі; у роботі [65] наведено дослідження відносно зміни віку населення.

Актуальними на сьогоднішній день є дослідження у галузі сталого розвитку та факторів, що впливають на сталий розвиток і медицину, оскільки вони дозволяють отримати картину розвитку для різних країн світу. Відновлення і прогнозування індикаторів сталого розвитку є також актуальним, тому що це дає можливість виконувати більш детальний аналіз стану соціального забезпечення населення у різних країнах.

3.2.1. Метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування

За рахунок аналізу даних про взаємозалежні часові ряди можна отримати додаткову інформацію про динаміку часового ряду [5]. Таку інформацію можна отримати шляхом побудови кореляційної матриці для змінних, стосовно яких відбувається дослідження [5].

Розглядається модель процесу випадкового блукання у просторі станів. Рівняння стану має такий вигляд [5]:

$$\mathbf{X}(i+1) = \mathbf{R} \times \mathbf{X}(i) + \mathbf{V}(i), \quad (3.9)$$

де $\mathbf{X}(i) = \begin{pmatrix} x_1(i) \\ \dots \\ x_m(i) \end{pmatrix}$ – вектор стану в i -й момент часу; m – кількість взаємозалежних

часових рядів; \mathbf{R} – кореляційна матриця; $\mathbf{V}(i) = \begin{pmatrix} v_1(i) \\ \dots \\ v_m(i) \end{pmatrix}$ – вектор шуму стану в i -й

момент часу з такими статистичними характеристиками [5]:

$$E[\mathbf{V}(i)] = \mathbf{q}^*, \quad E[(\mathbf{V}(i) - \mathbf{q}^*)(\mathbf{V}(k) - \mathbf{q}^*)^T] = \sigma^2 \times \delta(i - k),$$

де σ^2 – дисперсія шуму стану; $\delta(i - k)$ – значення дельта-функції Дірака в $(i - k)$ -й момент часу.

Кореляційна матриця \mathbf{R} має такий вигляд [24]:

$$\mathbf{R} = \begin{bmatrix} r_{x_1 x_1} & \dots & r_{x_m x_1} \\ \dots & \dots & \dots \\ r_{x_1 x_m} & \dots & r_{x_m x_m} \end{bmatrix}, \quad (3.10)$$

$$\text{де } r_{x_p x_m} = \frac{\frac{1}{N-1} \sum_{k=1}^N (x_p(k) - \bar{x}_p) \times (x_m(k) - \bar{x}_m)}{\sigma_{x_p} \times \sigma_{x_m}} - \text{елемент кореляційної матриці;}$$

$$\sigma_{x_p} = \sqrt{\frac{1}{N-1} \times \sum_{k=1}^N (x_p(k) - \bar{x}_p)^2} - \text{стандартне відхилення для } x_p;$$

$$\bar{x}_p = \frac{1}{N} \times \sum_{k=1}^N x_p(k) - \text{вибіркове середнє для } x_p; N - \text{кількість значень у вибірці даних.}$$

Значення вибірки даних вимірюються за умов наявності адитивного шуму, рівняння вимірів описується так [5]:

$$\mathbf{Y}(i) = \mathbf{X}(i) + \mathbf{W}(i), \quad (3.11)$$

$$\text{де } \mathbf{Y}(i) = \begin{bmatrix} y_1(i) \\ \dots \\ y_m(i) \end{bmatrix} - \text{вектор вимірів в } i\text{-й момент часу; } \mathbf{W}(i) = \begin{bmatrix} w_1(i) \\ \dots \\ w_m(i) \end{bmatrix} - \text{вектор}$$

шуму вимірів в i -й момент часу з такими статистичними характеристиками [5]:

$$E[\mathbf{W}(i)] = 0, \quad E[\mathbf{W}(i) \times \mathbf{W}(k)^T] = (\sigma^*)^2 \times \delta(i - k),$$

де $(\sigma^*)^2$ – дисперсія шуму вимірів. Вважаються апіорно невідомими статистичні характеристики σ та σ^* [5].

Метод відновлення пропущених значень для взаємозалежних часових рядів складається з наведених нижче етапів [5]:

1. Виконується обчислення елементів кореляційної матриці за формулою (3.10) для тих вимірів усіх взаємозалежних рядів, які не містять пропущені значення. Якщо пропущено від 1 до 3 перших значень ряду, то кореляційна матриця будується за всіма наступними відомими значеннями ряду для $i = l, \dots, N$, де $l \leq 3$. Якщо пропущено значення ряду, починаючи з $i = l$, де $l > 3$, то кореляційна матриця будується за значеннями $i = 1, \dots, l$. У даному методі враховується те, що взаємозалежні часові ряди мають однакову динаміку зміни значень. Тому якщо серед рядів, що розглядаються, є ряди, для яких значення коефіцієнтів кореляції є від'ємними, то ці ряди необхідно видалити з розгляду, оскільки від'ємні кореляційні коефіцієнти свідчать про протилежний напрямок зміни значень ряду. Також з розгляду необхідно видалити ряди, що показують слабку кореляцію по відношенню до ряду, що розглядається. При цьому значення порогу для слабкої кореляції визначається експертним шляхом, але рекомендується брати значення не менше, ніж 0,2. Якщо кореляційну матрицю побудувати неможливо, то пропущені значення замінюються останнім відомим значенням і відбувається перехід до кроку 3.

2. Розраховуються пропущені значення ряду за формулами (3.9) – (3.11).

3. Застосовується метод двобічного експоненційного згладжування до відновлених значень часового ряду [3]. Метод двобічного експоненційного згладжування складається з наведених нижче етапів [3]:

1) Виконується процедура експоненційного згладжування за формулою (1.13) для часового ряду зі значенням початкового наближення, що дорівнює першому значенню ряду. Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються, відповідно, за формулами (3.2) та (3.4) [3].

2) Виконується процедура згладжування у зворотному часі за формулою (3.5). Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються за формулами (3.2) та (3.4), відповідно, але при цьому в якості вхідних значень при обчисленні цих характеристик використовуються значення

S_k [3]. При обчисленні уточненої оцінки S_k^* ваги оцінок S_k надходять на вхід алгоритму у зворотному часі.

Таким чином, у розробленому методі відновлення пропущених значень для взаємозалежних часових рядів враховуються різні початкові умови для пропущених значень ряду під час побудови кореляційної матриці для взаємозалежних часових рядів, а також враховується випадок, коли кореляційну матрицю побудувати неможливо.

3.2.2. Приклади застосування методу відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування

У роботі [5] розглянуто вибірки даних про значення індексів кількості лікарняних ліжок, асфальтованих доріг, загальних значень прямих іноземних інвестицій, середньої кількості дітей, що навчаються у школі, рівня грамотності дорослих, державних витрат на освіту, сумарних державних витрат на освіту, військових витрат з 2004 по 2014 рік у Хорватії за даними Світового центру даних [66]. У роботі [5] виконано чотири різних експерименти з відновлення пропущених даних.

Відновлення пропущених значень виконується за допомогою таких методів [5]: розробленого методу відновлення для взаємозалежних часових рядів, що описується етапами 1–3 у розділі 3.2.1; заміни пропущених значень середніми арифметичними значеннями за формулою (1.12); за допомогою експоненційного згладжування за формулою (1.13).

Для оцінювання якості побудованих моделей у цілому обчислюються такі статистичні характеристики: коефіцієнт детермінації R^2 за формулою (3.1); сума квадратів похибок моделі SSE за формулою (3.2); статистика Дарбіна-Уотсона DW за формулою (3.3).

Приклад № 1

У вибірці даних про значення індексу середньої кількості дітей, що навчаються у школі, зроблено штучний пропуск значення за 2004 рік [5]. У роботі [5] розглядаються значення індексів середньої кількості дітей, що навчаються у школі, сумарних державних витрат на освіту з 2004 по 2014 рік у Хорватії за даними Світового центру даних [66] як взаємозалежні показники сталого розвитку.

У табл. 3.4 наведено значення статистичних характеристик, які показують якість моделей у цілому, розраховані за формулами (3.1), (3.2), (3.3) відповідно на проміжку з 2004 по 2014 рік [5]. У цій таблиці моделі позначено таким чином: заміна пропущених значень середніми арифметичними – СА; експоненційне згладжування – ЕЗ; метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування – МВЗВЧР.

Таблиця 3.4

Статистичні характеристики моделей

Статистична характеристика	СА	ЕЗ	МВЗВЧР
R^2	0,6349	0,6349	0,6459
SSE	8,6858	8,6858	8,5112
DW	2,0	2,0	2,0

Показано у табл. 3.4, що розроблений метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування показує перевагу перед методами відновлення середніми арифметичними значеннями та експоненційним згладжуванням за значенням коефіцієнта детермінації R^2 та суми квадратів похибок моделі SSE при відновленні першого значення часового ряду: значення коефіцієнта

детермінації R^2 для методу відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування є ближчим до ідеального значення 1 та більшим на 1,703%, ніж для середнього арифметичного та експоненційного згладжування; значення суми квадратів похибок моделі SSE для методу відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування є меншим на 2,01%, ніж для середнього арифметичного та експоненційного згладжування. Значення статистики Дарбіна-Уотсона DW є однаковими для усіх моделей.

Приклад № 2

У вибірці даних про значення індексу середньої кількості дітей, що навчаються у школі, виконано штучний пропуск значень за 2012, 2013 роки [5]. У роботі [5] розглянуто індекси значень кількості лікарняних ліжок, середньої кількості дітей, що навчаються у школі, сумарних державних витрат на освіту з 2004 по 2014 рік у Хорватії за даними Світового центру даних [66] як взаємозалежні показники сталого розвитку.

У табл. 3.5 наведено значення статистичних характеристик, які показують якість моделей у цілому, розраховані за формулами (3.1), (3.2), (3.3) відповідно на проміжку з 2004 по 2014 рік [5]. У цій таблиці моделі позначено аналогічно до табл. 3.4.

Таблиця 3.5

Статистичні характеристики моделей

Статистична характеристика	СА	ЕЗ	МВЗВЧР
R^2	0,9675	0,9562	0,9716
SSE	0,1589	0,2497	0,1487
DW	1	1,6036	1,7142

Результати, наведені у табл. 3.5, показують, що розроблений метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування показує перевагу перед іншими методами: значення коефіцієнта детермінації R^2 для методу відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування є ближчим до ідеального значення 1 та більшим на 0,422%, ніж для середнього арифметичного, більшим на 1,585%, ніж для експоненційного згладжування; значення суми квадратів похибок моделі SSE для розробленого методу є меншими на 6,419%, ніж для середнього арифметичного, є меншими на 40,449%, ніж для експоненційного згладжування; значення статистики Дарбіна-Уотсона DW для методу відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування є ближчими до ідеального значення 2 та більшими на 41,664%, ніж для середнього арифметичного, більшими на 6,452%, ніж для експоненційного згладжування.

Приклад № 3

Відновлюється фактично пропущене значення загальних іноземних інвестицій за 2004 рік [5]. У роботі [5] розглядаються значення кількості прямих загальних іноземних інвестицій, лікарняних ліжок з 2004 по 2014 рік за даними Світового центру даних [66] як взаємозалежні показники сталого розвитку. Відновлені значення, розраховані за наведеними вище методами відновлення пропущених даних, показано у табл. 3.6 [5]. У цій таблиці моделі позначено аналогічно до табл. 3.4.

Таблиця 3.6

Відновлені значення прямих загальних іноземних інвестицій

Рік	СА	ЕЗ	МВЗВЧР
2004	-1550927848	-1550927848	-1550341570

Однаковими є значення, відновлені за методами СА та ЕЗ, тому що відновлюється перше значення. Отримане значення за МВЗВЧР трохи більше значень, отриманих за методами СА та ЕЗ, при цьому реальне значення є невідомим.

Приклад № 4

Відновлюються фактично пропущені значення загального рівня грамотності дорослих за 2007 – 2010 роки [5]. У роботі [5] розглядаються тільки значення індексу загального рівня грамотності дорослих за 2004 – 2014 роки у Хорватії за даними Світового центру даних [66] без інших взаємозалежних часових рядів, оскільки у цьому випадку побудувати кореляційну матрицю для взаємозалежних рядів неможливо через однакові значення ряду за 2004 – 2006 роки.

Відновлені значення, розраховані за наведеними вище методами відновлення пропущених даних, описано у табл. 3.7, позначення зроблено аналогічно до табл. 3.4 [5].

Таблиця 3.7

Відновлені значення загального рівня грамотності дорослих

Рік	СА	ЕЗ	МВЗВЧР
2007	98,6360	98,1467	98,1475
2008	98,6360	98,1467	98,1493
2009	98,6360	98,1467	98,1555
2010	98,6360	98,1467	98,1760

Отримані за кожним з методів відновлення значення часового ряду є близькими, при цьому реальні значення є невідомими. Під час виконання моделювання у прикладі № 2 було показано, що розроблений метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування показав перевагу перед іншими методами, саме тому отримані за цим методом відновлені значення ряду рекомендовано використовувати.

3.2.3. Метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування

Метод прогнозування для взаємозалежних часових рядів складається з наведених нижче етапів [5]:

1. Виконується обчислення елементів кореляційної матриці за формулою (3.10) для $i = 1, \dots, N$. У даному методі передбачається, що взаємозалежні часові ряди мають однакову динаміку зміни значень. Тому якщо серед даних, які розглядаються, є ряди, для яких значення коефіцієнтів кореляції є від'ємними, то ці ряди необхідно прибрати з розгляду, оскільки від'ємні кореляційні коефіцієнти свідчать про протилежний напрямок зміни значень ряду. Також з розгляду необхідно прибрати ряди, які показують слабку кореляцію по відношенню до ряду, що розглядається. При цьому значення порогу для слабкої кореляції визначається експертним шляхом, але рекомендується обирати значення не менше, ніж 0,2. Якщо кореляційну матрицю неможливо побудувати, то пропущені значення замінюються на останнє відоме значення та виконується перехід до кроку 3.

2. Розраховуються прогнозні значення ряду за формулами (3.9) – (3.11).

3. До прогнозних значень ряду застосовується метод двобічного експоненційного згладжування, що описується етапами 1 – 3 у розділі 3.2.1.

Таким чином, у розробленому методі прогнозування для взаємозалежних часових рядів враховуються можливості аналізу для підтвердження

взаємозалежності часових рядів після побудови кореляційної матриці та враховується випадок, коли кореляційну матрицю побудувати неможливо.

3.2.4. Приклад застосування методу прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування

У роботі [5] на основі вибірок даних про індекси значень середньої кількості дітей, що навчаються у школі, загальних державних витрат на освіту з 2004 по 2013 рік у Хорватії за даними Світового центру даних [66] виконується прогнозування значень ряду індексу середньої кількості дітей, що навчаються у школі, за 2013, 2014 роки.

Виконується прогнозування за допомогою розробленого методу прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування та експоненційного згладжування [5].

Після побудови математичних моделей обчислено значення прогнозних характеристик, а саме, значення середньоквадратичної похибки RMSE, середньої відсоткової абсолютної похибки MAPE та коефіцієнта Тейла U за формулами (2.22), (2.23), (2.24) відповідно. Критерієм покращення придатності побудованої моделі для прогнозування є мінімізація перерахованих вище прогнозних характеристик.

У табл. 3.8 показано значення цих прогнозних характеристик, розрахованих під час прогнозування значень ряду індексу середньої кількості дітей, що навчаються у школі з використанням наведених вище методів [5]. У цій таблиці методи позначено таким чином: експоненційне згладжування – ЕЗ; метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування – МПВЧР.

Таблиця 3.8

Значення прогнозних характеристик моделей

Прогнозна характеристика	ЕЗ	МПВЧР
RMSE	0,0319	0,0212
MAPE	0,0264	0,0172
U	0,0002	0,0001

Результати, наведені у табл. 3.8, показують, що розроблений метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування показує перевагу перед методом експоненційного згладжування за значеннями усіх прогнозних характеристик: значення середньоквадратичної похибки RMSE є меншим для розробленого методу на 33,542%, ніж для експоненційного згладжування; значення середньої відсоткової абсолютної похибки MAPE є меншим для розробленого методу на 34,848%, ніж для експоненційного згладжування; значення коефіцієнта Тейла U є меншим на 50%, ніж для експоненційного згладжування.

3.3. Розробка методу ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів

За допомогою алгоритмів оптимальної фільтрації неможливо розв'язати певний особливий клас задач. До таких задач відносяться задачі обробки даних про сонячну активність, задачі океанографічних досліджень. Ці задачі можна розглядати як задачі з уточнення минулих по відношенню до теперішнього часу параметрів, оскільки вони мають певні особливості. За допомогою методу згладжування можуть вирішуватися такі задачі [4]. Відновлення істинних прихованих закономірностей динамічних процесів при обробці експериментальних даних, що описують сонячну активність, таких, як числа

Вольфа та сонячні радіофлюкси при довжині хвилі 10,7 см, є актуальною задачею на сьогоднішній день [4]. Побудова моделі за даними, що описують сонячну активність, є складною задачею, тому що фізичні процеси, які відбуваються на Сонці, є недостатньо вивченими та об'єм даних є обмеженим.

Для правильної попередньої обробки даних та подальшого прогнозування є важливими на сьогоднішній день подальші дослідження стосовно створення нових методів для відновлення істинних закономірностей, відновлення пропущених значень та прогнозування даних про сонячну активність.

Різні групи науковців займались дослідженням можливостей відновлення істинних закономірностей та прогнозування сонячних даних. Розглянуто роботи, за якими можна прослідкувати розвиток наукової думки відносно цього питання.

У роботі [28] Хетеуей та інші запропонували рівняння для опису моделі 11-річного сонячного циклу. Пізніше ці автори у статті [27] запропонували метод розрахунку 13-місячного ковзного середнього для обробки даних про сонячну активність. Цей метод є розповсюдженим на сьогоднішній день. Сальвадор запропонував математичну модель для циклу чисел Вольфа з використанням припливно-обертальної теорії Ієна Вілсона у роботі [67]. Виконано дослідження похибок методу 13-місячного ковзного середнього та визначено умови, за яких застосування цього методу при аналізі сонячної активності викривляє значні особливості цього процесу у статті [68].

Багато науковців виконували дослідження проблеми прогнозування сонячних даних. У роботі [28] Хетеуей та інші показують, що запропоноване рівняння для опису моделі 11-річного сонячного циклу може бути застосованим для прогнозування. У роботі [27] виконано огляд багатьох існуючих методів прогнозування сонячних даних, всі методи прогнозування поділено на регресійні та методи, що базуються на аналізі попередніх даних. Байєсівський підхід для прогнозування сонячних циклів з використанням рівняння Фокера-Планка запропоновано Ноблем та Вітландом у статті [69]. У роботі [67]

показано, що наведену модель можна використати для опису закономірностей та прогнозування чисел Вольфа.

Розглянуто багато методів для відновлення закономірностей та прогнозування сонячних даних у наведених вище роботах. Також у цих роботах наведено дослідження, що є критичними та показують недоліки вже розроблених методів за певних умов. Тому задача розробки нових методів для відновлення істинних закономірностей, пропущених значень та прогнозування сонячних даних є актуальною на сьогоднішній день.

3.3.1. Метод ковзного двобічного експоненційного згладжування

У роботі [11] запропоновано метод ковзного двобічного експоненційного згладжування, що полягає у побудові цього методу на інтервалі 13, щоб працювати з даними у такому ж діапазоні, як і для 13-місячного ковзного середнього. Цей метод є розвитком та вдосконаленням квазіоптимальної процедури згладжування, наведеної у роботі [60].

Перша частина процедури ковзного двобічного тринадцятимісячного експоненційного згладжування полягає в отриманні оцінок за допомогою традиційного методу експоненційного згладжування [11]. При цьому початкове наближення обирається рівним першому значенню ряду з відповідного інтервалу на кожному ковзному тринадцятимісячному інтервалі [11]. Ця процедура описується такою формулою [11]:

$$S_k = \alpha \times y(k) + (1 - \alpha) \times S_{k-1}, \quad (3.12)$$

де S_k – значення експоненційного середнього в k -й момент часу; α – коефіцієнт згладжування; $y(k)$ – значення ряду в k -й момент часу; S_{k-1} – значення експоненційного середнього в $(k-1)$ -й момент часу, періоди часу $k = 1, \dots, 13, k = 2, \dots, 14, \dots, k = N - 12, \dots, N$.

У роботі [4] в якості критерію вибору оптимального значення параметра згладжування використовується мінімум значень суми квадратів похибок SSE

та інтегрального критерію I , що обчислюються, відповідно, за формулами (3.2) та (3.4).

Друга частина процедури ковзного двобічного експоненційного згладжування полягає у застосуванні до отриманих оцінок процедури згладжування у зворотному часі на кожному інтервалі $k = N - 1, \dots, N - 12, k = N - 2, \dots, N - 13, \dots, k = 12, \dots, 1$ [11]. Кожне уточнене значення S_k^* розраховується так [11]:

$$S_k^* = \alpha \times S_k + (1 - \alpha) \times S_{k+1}^*, \quad (3.13)$$

де S_k^* – значення експоненційного середнього у зворотному часі в k -й момент часу; α – коефіцієнт згладжування; S_k – значення експоненційного середнього в k -й момент часу; S_{k+1}^* – значення експоненційного середнього у зворотному часі в $(k+1)$ -й момент часу, тут значення $k = N - 1, \dots, N - 12, \dots, k = 12, \dots, 1$, початкові значення $S_N^* = S_N, \dots, S_{13}^* = S_{13}$.

Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються за формулами (3.2) та (3.4), відповідно, але при цьому в якості вхідних значень при обчисленні цих характеристик використовуються значення S_k [4]. При обчисленні уточненої оцінки S_k^* ваги оцінок S_k надходять на вхід алгоритму у зворотному часі [4].

У роботі [11] описано, що після цього в якості оцінки значення вибірки даних використовується отримане значення у середній точці з кожного інтервалу, тобто на інтервалі $k = 1, \dots, 13$ взято значення S_7^* , на інтервалі $k = 2, \dots, 14$ взято значення S_8^* і т. д. Оцінки для перших шести та останніх шести значень вибірки даних взято як відповідні значення експоненційного середнього у зворотному часі на першому та останньому інтервалах оцінювання відповідно, тобто S_1^*, \dots, S_6^* на інтервалі $k = 1, \dots, 13$ та S_{N-5}^*, \dots, S_N^* на інтервалі $k = N - 12, \dots, N$ [11].

Для прогнозування часових рядів метод ковзного двобічного експоненційного згладжування запропоновано застосовувати так [4]: необхідно використовувати повну описану процедуру і розрахунки проводяться за формулами (3.12) та (3.13), при цьому значення параметра згладжування для згладжування у прямому часі використовується таке ж, як і для всіх відомих значень, в якості початкового значення використовується останнє відоме значення.

Метод ковзного двобічного експоненційного згладжування запропоновано застосовувати для відновлення істинних закономірностей часових рядів шляхом застосування цього методу до значень на часовому інтервалі, на якому задані ці значення, відповідно до процедури, що описується рівняннями (3.12) та (3.13). Для знаходження параметра згладжування на кожному з етапів двобічного експоненційного згладжування застосовується критерій мінімізації значень, що описуються рівняннями (3.2) та (3.4).

Для виконання подальшої обробки ряду, зокрема, прогнозування, важливим є відновлення пропущених значень часових рядів з отриманням точних оцінок значень замість пропущених. У роботі [4] з використанням методу ковзного двобічного експоненційного згладжування розроблено алгоритм для відновлення пропущених значень ряду.

Алгоритм № 1 описується так [4]:

1. Визначається параметр згладжування. Якщо перед пропуском знаходиться від 1 до 7 значень ряду, то для знаходження параметра згладжування на кожному з етапів ковзного двобічного експоненційного згладжування за критерієм мінімізації значень, що описуються рівняннями (3.2) та (3.4), запропоновано брати значення ряду, що знаходяться після пропущеного значення. Розглядається в якості граничного кількісного значення 7, тому що при застосуванні методу ковзного двобічного експоненційного згладжування на першому інтервалі обирається сьома оцінка. Якщо перед пропуском знаходиться більше семи значень ряду, то для знаходження параметра згладжування на кожному з етапів ковзного двобічного

експоненційного згладжування за критерієм мінімізації значень, що описуються рівняннями (3.2) та (3.4), запропоновано брати значення ряду, що знаходяться перед пропущеним значенням.

2. Зі значенням початкового наближення, рівним середньому арифметичному серед усіх відомих значень, та з використанням в якості відомого значення виміру останнього значення перед пропущеним значенням запропоновано застосовувати процедуру ковзного двобічного експоненційного згладжування, що описується рівняннями (3.12) та (3.13) відповідно.

Таким чином, метод ковзного двобічного експоненційного згладжування може бути ефективно застосованим для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.

3.3.2. Приклади застосування методу ковзного двобічного експоненційного згладжування для відновлення пропущених значень та істинних закономірностей часових рядів

У роботі [4] для вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см (далі використовується назва індекс F10.7) в одиницях виміру SFU з січня 2013 року до березня 2016 року (39 значень) за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70] виконується дослідження стосовно штучного створення пропусків та відновлення пропущених значень.

За допомогою таких методів виконується відновлення пропущених значень індексу F10.7 [4]: методу ковзного двобічного експоненційного згладжування за формулами (3.12), (3.13), (3.2), (3.4) за алгоритмом № 1; методу експоненційного згладжування за формулами (1.13) та (3.2), (3.4). Для оцінювання якості побудованих моделей у цілому при відновленні пропущених значень обчислюються статистичні характеристики за формулами (3.1), (3.2), (3.3).

Випадок № 1

У вибірці даних щодо індексу F10.7 зроблено штучний пропуск 3 значень з номерами $N = 7, 8, 9$ та виконано відновлення за двома описаними вище методами [4].

У табл. 3.9 наведено статистичні характеристики, що показують якість побудованих моделей у цілому для випадку № 1 при відновленні значень індексу F10.7, розраховані за формулами (3.1), (3.2), (3.3) на проміжку $N=1,...,39$ [4]. У цій таблиці моделі позначено таким чином: експоненційне згладжування – ЕЗ та ковзне двобічне експоненційне згладжування – КДЕЗ.

Таблиця 3.9

Статистичні характеристики моделей

Статистична характеристика	ЕЗ	КДЕЗ
R^2	0,9477	0,9565
SSE	103,3539	100,1215
DW	1,3125	1,3372

За наведеними у табл. 3.9 значеннями статистичних характеристик можна зробити такі висновки: значення коефіцієнта детермінації R^2 для ковзного двобічного експоненційного згладжування є ближчим до ідеального значення 1 та більшим на 0,92%, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для розробленого методу є меншою на 3,128%, ніж для експоненційного згладжування; статистика Дарбіна-Уотсона DW для ковзного двобічного експоненційного згладжування є ближчою до ідеального значення 2 та більшою на 1,847%, ніж для експоненційного згладжування.

Випадок № 2

У вибірці даних щодо індексу F10.7 зроблено штучний пропуск 3 значень з номерами $N = 30, 31, 32$ та виконано відновлення за двома описаними вище методами [4].

У табл. 3.10 наведено статистичні характеристики моделей аналогічно до табл. 3.9 для випадку № 2 при відновленні значень індексу F10.7 [4].

Таблиця 3.10

Статистичні характеристики моделей

Статистична характеристика	ЕЗ	КДЕЗ
R^2	0,9451	0,9975
SSE	274,1465	95,3291
DW	1,2370	1,9452

За наведеними у табл. 3.10 значеннями статистичних характеристик можна зробити такі висновки: значення коефіцієнта детермінації R^2 для ковзного двобічного експоненційного згладжування є ближчим до ідеального значення 1 та більшим на 5,253%, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для розробленого методу є меншою на 65,227%, ніж для експоненційного згладжування; статистика Дарбіна-Уотсона DW для ковзного двобічного експоненційного згладжування є ближчою до ідеального значення 2 та більшою на 36,408%, ніж для експоненційного згладжування.

За наведеними у табл. 3.9 та табл. 3.10 значеннями статистичних характеристик метод ковзного двобічного експоненційного згладжування показує перевагу перед методом експоненційного згладжування при відновленні пропущених значень індексу F10.7. Тому розроблений метод є ефективним для відновлення пропущених значень часових рядів за різних початкових умов.

На основі цієї ж вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см відновлюються істинні значення ряду [4]. До цих даних застосовано метод ковзного двобічного експоненційного згладжування за формулами (3.12), (3.13), (3.2), (3.4) та метод 13-місячного ковзного середнього за формулою (1.15) [4]. Метод 13-місячного ковзного середнього традиційно використовується для аналізу варіацій сонячної активності [27]. Проте цей метод може викривляти величини тривалості сонячного циклу, максимуму та мінімуму, що показано у роботі [27], оскільки він не відфільтровує високочастотні коливання з періодом менше одного року. Порівняння отриманих результатів виконано за допомогою критерію близькості, що полягає у пошуку мінімуму для виразу (3.2) [4].

Виміряні значення індексу F10.7 та результати, отримані за кожною з моделей при відновленні істинних значень ряду, показано на рис. 3.2 [4].

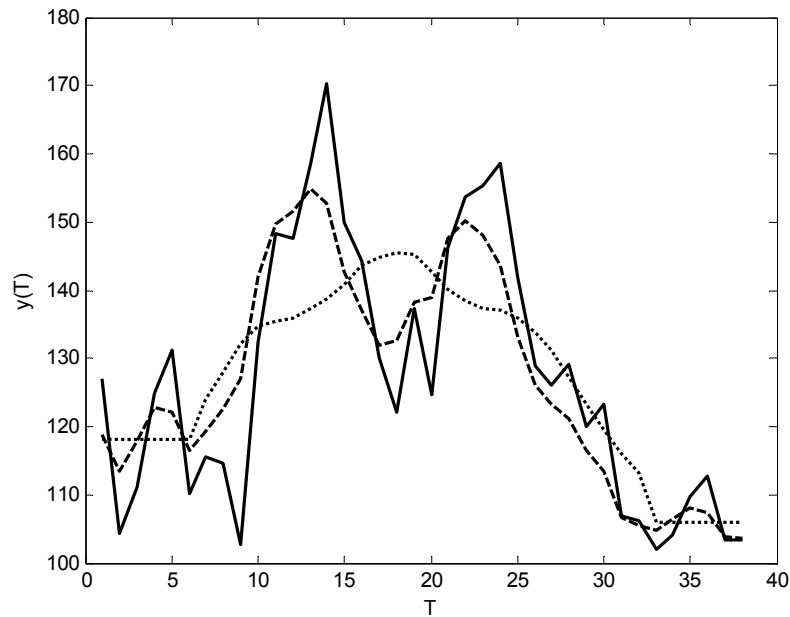


Рис. 3.2. Значення $y(i)$ та $\hat{y}(i)$ для вибірки значень індексу F10.7, де $-$ – виміряні значення $y(i)$; $- -$ – оцінки $\hat{y}(i)$, отримані за методом ковзного двобічного експоненційного згладжування; \cdots – оцінки $\hat{y}(i)$, отримані за методом 13-місячного ковзного середнього

У роботі [71] Дж.М. Херст розглянув вплив ширини вікна на зміну характеристик точності ковзного середнього та показав, що ковзне середнє може значно викривити процес, знизити короткочасні коливання, у той час, як коливання, період яких значно перевищує ширину вікна, відновлюються майже без викривлень. На рис. 3.2 показано, що на часовому інтервалі $[10, 25]$ для моделі, побудованої за методом 13-місячного ковзного середнього, відбувається суттєве викривлення процесу: на часових інтервалах $[10, 15]$ та $[20, 25]$ коливання обнуляються, на інтервалі $[15, 20]$ відбувається обертання коливань індексу F10.7 [4]. На цьому часовому інтервалі ковзне двобічне експоненційне згладжування не викривляє коливання під час відновлення істинних закономірностей індексу F10.7 [4].

У табл. 3.11 наведено значення критерію близькості, що розраховується за формулою (3.2) [4]. У цій таблиці моделі позначено таким чином: ковзне двобічне експоненційне згладжування – КДЕЗ; 13-місячне ковзне середнє – ТКС.

Таблиця 3.11

Значення критеріїв близькості для моделей

Назва критерію	КДЕЗ	ТКС
SSE	2441,2	6115,3

За допомогою застосування критерію близькості до значень, наведених у табл. 3.11, можна зробити висновок, що метод ковзного двобічного експоненційного згладжування показує перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей індексу F10.7, тому що значення SSE для методу ковзного двобічного експоненційного згладжування є меншим на 60,08%, ніж для методу 13-місячного ковзного середнього.

3.3.3. Приклад застосування методу ковзного двобічного експоненційного згладжування для прогнозування часових рядів

У роботі [4] на основі вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU з січня 2013 року до грудня 2015 року (36 значень) [70] виконується прогнозування значень на наступні 3 місяці. Прогнозування виконується за допомогою таких методів [4]: методу ковзного двобічного експоненційного згладжування за формулами (3.12), (3.13), (3.2), (3.4) та методу, запропонованого у роботі [28], що описується формулами (1.16) – (1.18).

У табл. 3.12 наведено прогнозні характеристики, обчислені за формулами (2.22), (2.23), (2.24) відповідно на проміжку $N=37, \dots, 39$ [4]. У цій таблиці

моделі позначено таким чином: експоненційний підхід, що описується формулами (1.16)– (1.18) – ЕП та ковзне двобічне експоненційне згладжування– КДЕЗ.

Таблиця 3.12

Значення прогнозних характеристик для моделей

Прогнозна характеристика	ЕП	КДЕЗ
RMSE	20,1355	18,1569
MAPE	8,8084	7,6241
U	0,0065	0,0051

Метод ковзного двобічного експоненційного згладжування показує перевагу перед традиційним експоненційним підходом при прогнозуванні за значеннями прогнозних характеристик, наведених у табл. 3.12: значення середньоквадратичної похибки RMSE є меншим для розробленого методу на 9,826%, ніж для експоненційного згладжування; значення середньої відсоткової абсолютної похибки MAPE є меншим для методу ковзного двобічного експоненційного згладжування на 13,445%, ніж для експоненційного згладжування; значення коефіцієнта Тейла U є меншим для розробленого методу на 21,538%, ніж для експоненційного згладжування.

3.4. Розробка методу адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів

Аналіз, математичне моделювання і прогнозування нелінійних нестационарних процесів, зокрема даних, що описують сонячні цикли, є актуальною сучасною задачею обробки даних. Відомою роботою, в якій описується модель 11-річного сонячного циклу, є робота Хетеуея та інших [28].

Пізніше ці автори у роботі [27] запропонували метод розрахунку 13-місячного ковзного середнього для обробки даних про сонячну активність. На сьогоднішній день цей метод є широко розповсюдженим. У роботах [72-74] виконуються дослідження стосовно динаміки сонячного циклу. Ефект Вальдмайєра є специфічним тільки для чисел Вольфа, що показано у роботі [75]. У роботах [76-79] наведено історичний розвиток дослідження сонячних плям, перехід від чисел Вольфа до Міжнародного індексу сонячних плям. Також важливим є розв'язання проблеми невизначеності та знаходження уточнених оцінок для числа сонячних плям, ці дослідження виконано у роботах [80-83]. Математичну модель для циклу чисел Вольфа з використанням припливно-обертальної теорії Ієна Вілсона запропоновано Сальвадором у роботі [67].

Актуальною на сьогоднішній день є проблема прогнозування сонячних даних. У роботі [28] Хетеуей та інші показують, що запропоноване рівняння для опису 11-річного сонячного циклу може бути застосованим для прогнозування. У роботах [84-86] наведено дослідження стосовно використання даних про динаміку зміни сонячних радіофлюксів для прогнозування сонячних циклів. У роботі [27] зроблено докладний огляд багатьох існуючих методів прогнозування сонячних даних. Байєсівський підхід до прогнозування сонячних циклів з використанням рівняння Фокера-Планка запропоновано Ноблем та Вітландом у статті [69]. У роботі [67] показано, що запропоновану модель можна використати як для опису закономірностей, так і для прогнозування чисел Вольфа. У роботі [87] показано на сьогоднішній день стан досліджень, що стосуються прогнозування сонячних циклів, та запропоновано підхід для підвищення точності прогнозів.

Актуальними є дослідження стосовно відновлення істинних закономірностей та пропущених значень сонячних даних, тому що знаходження більш точних оцінок для їх використання у подальшій обробці та прогнозуванні сонячних даних є важливим. Актуальною проблемою є також прогнозування

сонячних даних, тому що знаходження більш точних прогнозних оцінок є важливим для дослідження сонячних даних.

3.4.1. Метод адаптивного ковзного двобічного експоненційного згладжування

Метод ковзного двобічного експоненційного згладжування запропоновано у роботі [11]. У роботі [11] цей метод застосовується на інтервалі 13, щоб працювати з даними у такому ж діапазоні, як і для 13-місячного ковзного середнього. Метод ковзного двобічного експоненційного згладжування є розвитком квазіоптимальної процедури згладжування, що описана у роботі [60].

Запропонований у роботі [11] метод ковзного двобічного експоненційного згладжування застосовується на інтервалі 13 місяців, при цьому вибір такого інтервалу здійснено на основі експертних даних. Проте на практиці значення оптимального інтервалу для цього методу може відрізнитись від значення 13. Тому у роботі [6] запропоновано адаптивний метод ковзного двобічного експоненційного згладжування, в якому виконується визначення змінного інтервалу та параметра згладжування.

Перший етап методу адаптивного ковзного двобічного експоненційного згладжування полягає в отриманні оцінок за допомогою традиційного методу експоненційного згладжування [6]. Початкове наближення обирається рівним першому значенню ряду з відповідного інтервалу на кожному ковзному інтервалі [6]. Ця процедура описується такою формулою [6]:

$$S_k = \alpha \times y(k) + (1 - \alpha) \times S_{k-1}, \quad (3.14)$$

де S_k – значення експоненційного середнього в k -й момент часу; α – коефіцієнт згладжування; $y(k)$ – значення ряду в k -й момент часу; S_{k-1} – значення експоненційного середнього в $(k-1)$ -й момент часу, періоди часу

$k = 1, \dots, M, k = 2, \dots, M + 1, \dots, k = N - M + 1, \dots, N$; M – значення ковзного інтервалу.

У даному методі пропонується обирати ковзний інтервал серед таких значень: 3, 5, 7, 9, 11, 13 [6]. У роботі [6] в якості критерію вибору оптимального значення параметра згладжування використовується мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються, відповідно, за формулами (3.2) та (3.4). Критерії вибору оптимального ковзного інтервалу аналогічні до критеріїв вибору оптимального значення коефіцієнта згладжування [6]. Виконується одночасне оцінювання коефіцієнта згладжування та ковзного інтервалу [6].

Другий етап методу адаптивного ковзного двобічного експоненційного згладжування полягає у застосуванні до отриманих оцінок процедури згладжування у зворотному часі на кожному інтервалі $k = N - 1, \dots, N - M + 1, k = N - 2, \dots, N - M, \dots, k = M - 1, \dots, 1$ [6]. Кожне уточнене значення S_k^* розраховується так [6]:

$$S_k^* = \alpha \times S_k + (1 - \alpha) \times S_{k+1}^*, \quad (3.15)$$

де S_k^* – значення експоненційного середнього у зворотному часі в k -й момент часу; α – коефіцієнт згладжування; S_k – значення експоненційного середнього в k -й момент часу; S_{k+1}^* – значення експоненційного середнього у зворотному часі в $(k + 1)$ -й момент часу, тут значення $k = N - 1, \dots, N - M + 1, \dots, k = M - 1, \dots, 1$, початкові значення $S_N^* = S_N, \dots, S_M^* = S_M$.

Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюються, відповідно, за формулами (3.2) та (3.4), але при цьому в якості вхідних значень при обчисленні цих характеристик використовуються значення S_k [6]. При згладжуванні у зворотному часі ковзний інтервал обирається таким самим, як і при згладжуванні у прямому часі [6]. При обчисленні уточненої оцінки S_k^* ваги оцінок S_k надходять на вхід алгоритму у зворотному часі [6].

Після цього в якості оцінки значення вибірки даних використовується отримане значення у середній точці з кожного інтервалу, тобто на інтервалі $k = 1, \dots, M$ взято значення $S_{\frac{M+1}{2}}^*$, на інтервалі $k = 2, \dots, M + 1$ взято значення

$S_{\frac{M+2}{2}}^*$ і т. д. [6]. Оцінки для перших $\frac{M+1}{2}$ та останніх $\frac{M+1}{2}$ значень вибірки

даних взято як відповідні значення експоненційного середнього у зворотному часі на першому та останньому інтервалах оцінювання відповідно, тобто $S_1^*, \dots, S_{\frac{M-1}{2}}^*$ на інтервалі $k = 1, \dots, M$ та $S_{\frac{2N-M+3}{2}}^*, \dots, S_N^*$ на інтервалі $k = N - M + 1, \dots, N$ [6].

Таким чином, розроблений метод адаптивного ковзного двобічного експоненційного згладжування може бути ефективно застосованим на практиці, оскільки містить у собі процедуру вибору оптимального ковзного інтервалу та застосовується критерій вибору оптимального значення параметра згладжування.

3.4.2. Застосування методу адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів

Для прогнозування значень часових рядів метод адаптивного ковзного двобічного експоненційного згладжування запропоновано застосовувати таким чином [6]: необхідно використовувати повну описану процедуру і розрахунки проводяться за формулами (3.14) та (3.15), при цьому значення параметра згладжування та ковзний інтервал використовуються такі ж, як і для всіх відомих значень, в якості початкового значення використовується останнє відоме значення.

Метод адаптивного ковзного двобічного експоненційного згладжування може бути застосованим для відновлення істинних закономірностей часових

рядів шляхом застосування цього методу до значень на часовому інтервалі, на якому задано ці значення, відповідно до процедури, що описується рівняннями (3.14) та (3.15). Для знаходження коефіцієнта згладжування та ковзного інтервалу на кожному з етапів адаптивного ковзного двобічного експоненційного згладжування застосовуються критерії мінімізації значень, що описуються рівняннями (3.2) та (3.4).

На сьогоднішній день важливим є відновлення пропущених значень часових рядів з отриманням точних оцінок значень замість пропущених для виконання подальшої обробки ряду, наприклад, для прогнозування. У роботі [6] з використанням методу адаптивного ковзного двобічного експоненційного згладжування запропоновано алгоритм для відновлення пропущених значень ряду.

Алгоритм № 1 [6]:

1. Визначається параметр згладжування та ковзний інтервал. Якщо перед пропуском знаходиться від 1 до $\frac{M+1}{2}$ значень ряду, то для знаходження параметра згладжування та ковзного інтервалу на кожному з етапів адаптивного ковзного двобічного експоненційного згладжування за критерієм мінімізації значень, що описуються рівняннями (3.2) та (3.4), запропоновано брати значення ряду, що знаходяться після пропущеного значення. Розглядається в якості граничного кількісного значення $\frac{M+1}{2}$, тому що при застосуванні методу адаптивного ковзного двобічного експоненційного згладжування на першому інтервалі обирається $\frac{M+1}{2}$ оцінка. Якщо перед пропуском знаходиться більше $\frac{M+1}{2}$ значень ряду, то для знаходження параметра згладжування та ковзного інтервалу на кожному з етапів адаптивного ковзного двобічного експоненційного згладжування за критерієм

мінімізації значень, що описуються рівняннями (3.2) та (3.4), запропоновано використовувати значення ряду, що знаходяться перед пропущеним значенням.

2. Зі значенням початкового наближення, рівним середньому арифметичному серед усіх відомих значень, та з використанням в якості відомого значення виміру останнього значення перед пропущеним значенням запропоновано застосовувати процедуру адаптивного ковзного двобічного експоненційного згладжування, що описується рівняннями (3.14) та (3.15) відповідно.

Таким чином, метод адаптивного ковзного двобічного експоненційного згладжування може бути ефективно застосованим для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.

3.4.3. Приклади відновлення пропущених значень та істинних закономірностей часових рядів з використанням методу адаптивного ковзного двобічного експоненційного згладжування

У роботі [6] на основі вибірки даних про місячні значення чисел Вольфа з січня 2013 року до серпня 2016 року (44 значення) за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70] відновлюються істинні значення ряду. До цих даних застосовано метод адаптивного ковзного двобічного експоненційного згладжування за формулами (3.14), (3.15), (3.2), (3.4) та метод 13-місячного ковзного середнього, що описується формулою (1.15) [6]. Критерій близькості, що полягає у пошуку мінімуму для виразу (3.2), використано для порівняння отриманих результатів [6].

На рис. 3.3 показано виміряні значення чисел Вольфа та результати, отримані за кожною з моделей при відновленні істинних значень ряду [6].

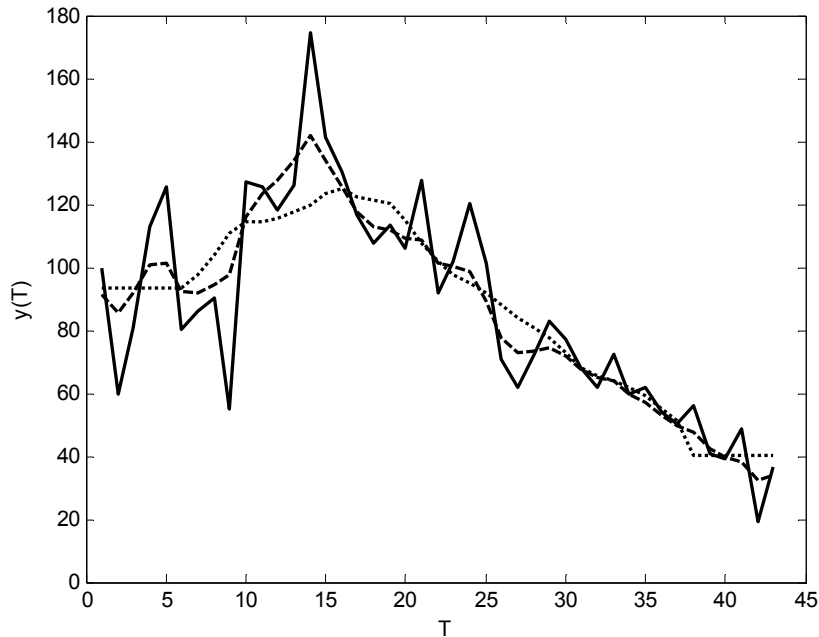


Рис. 3.3. Значення $y(i)$ та $\hat{y}(i)$ для вибірки значень чисел Вольфа, де $-$ – виміряні значення $y(i)$; $- -$ – оцінки $\hat{y}(i)$, отримані за методом адаптивного ковзного двобічного експоненційного згладжування; \cdots – оцінки $\hat{y}(i)$, отримані за методом 13-місячного ковзного середнього

У табл. 3.13 наведено значення критерію близькості, що розраховується за формулою (3.2) [6]. У цій таблиці моделі позначено таким чином: адаптивне ковзне двобічне експоненційне згладжування – АКДЕЗ; 13-місячне ковзне середнє – ТКС.

Таблиця 3.13

Значення критеріїв близькості для моделей

Назва критерію	АКДЕЗ	ТКС
SSE	6857,8	13459,1

У роботі [6] на основі вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см (далі використовується назва індекс

F10.7) в одиницях виміру SFU з січня 2013 року до серпня 2016 року (44 значення) за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70] відновлюються істинні значення ряду. До цих даних застосовано метод адаптивного ковзного двобічного експоненційного згладжування за формулами (3.14), (3.15), (3.2), (3.4) та метод 13-місячного ковзного середнього, що описується формулою (1.15) [6]. Для порівняння отриманих результатів використано критерій близькості, що полягає у пошуку мінімуму для виразу (3.2) [6].

На рис. 3.4, наведеному у роботі [6], показано виміряні значення індексу F10.7 та результати, отримані за кожною з моделей при відновленні істинних значень ряду.

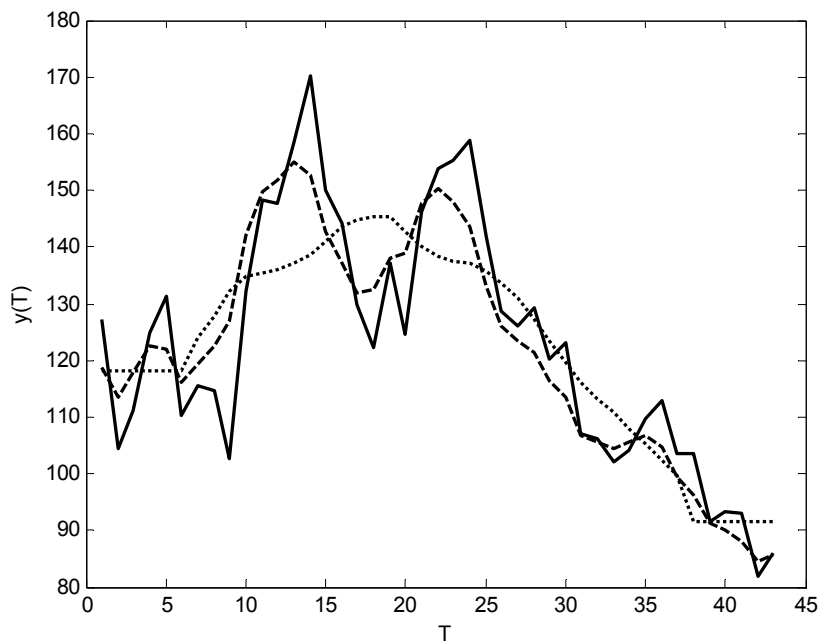


Рис. 3.4. Значення $y(i)$ та $\hat{y}(i)$ для вибірки значень індексу F10.7, де $-$ – виміряні значення $y(i)$; $- -$ – оцінки $\hat{y}(i)$, отримані за методом адаптивного ковзного двобічного експоненційного згладжування; \cdots – оцінки $\hat{y}(i)$, отримані за методом 13-місячного ковзного середнього

У табл. 3.14 наведено значення критерію близькості, що розраховується за формулою (3.2) [6]. У цій таблиці моделі позначено таким чином: адаптивне ковзне двобічне експоненційне згладжування – АКДЕЗ; 13-місячне ковзне середнє – ТКС.

Таблиця 3.14

Значення критеріїв близькості для моделей

Назва критерію	АКДЕЗ	ТКС
SSE	2593,4	6528,3

Дж.М. Херст у роботі [71] розглянув вплив ширини вікна на зміну характеристик точності ковзного середнього і показав, що ковзне середнє може значно викривити процес, знизити короточасні коливання, у той час, як коливання, період яких значно перевищує ширину вікна, відновлюються майже без викривлень. На рис. 3.3 показано, що на часових інтервалах [5, 10] та [15, 20] відбувається викривлення процесу для моделі, побудованої за методом 13-місячного ковзного середнього: відбувається обертання коливань чисел Вольфа [6]. На рис. 3.4 показано, що на часовому інтервалі [10, 25] для моделі, побудованої за методом 13-місячного ковзного середнього, відбувається суттєве викривлення процесу: на часових інтервалах [10, 15] та [20, 25] коливання обнуляються, на інтервалі [15, 20] відбувається обертання коливань індексу F10.7 [6]. У той же час, адаптивне ковзне двобічне експоненційне згладжування не викривляє коливання під час відновлення істинних закономірностей індексу F10.7 [6].

Наведене у табл. 3.13 значення SSE для методу адаптивного ковзного двобічного експоненційного згладжування є меншим на 49,047%, ніж для методу 13-місячного ковзного середнього. Наведене у табл. 3.14 значення SSE для методу адаптивного ковзного двобічного експоненційного згладжування є меншим на 60,274%, ніж для методу 13-місячного ковзного середнього. Застосовуючи критерій близькості до значень, наведених у табл. 3.13 та

табл. 3.14, можна зробити висновок, що метод адаптивного ковзного двобічного експоненційного згладжування показує перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей чисел Вольфа та індексу F10.7.

У роботі [6] на основі такої ж вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70] виконується дослідження стосовно штучного створення пропусків та відновлення пропущених значень.

Відновлення пропущених значень індексу F10.7 виконується за допомогою таких методів [6]: методу адаптивного ковзного двобічного експоненційного згладжування за формулами (3.14), (3.15), (3.2), (3.4) за алгоритмом № 1; методу експоненційного згладжування за формулами (1.13) та (3.2), (3.4). Для оцінювання якості побудованих моделей у цілому при відновленні пропущених значень обчислюються статистичні характеристики за формулами (3.1), (3.2), (3.3). Встановлено, що значення ковзного інтервалу дорівнює 11 для усіх випадків відновлення [6].

Випадок № 1

У вибірці даних щодо індексу F10.7 зроблено штучний пропуск 3 значень з номерами $N = 6, 7, 8$ та виконано відновлення за описаними вище методами [6].

Випадок № 2

У вибірці даних щодо індексу F10.7 зроблено штучний пропуск 3 значень з номерами $N = 28, 29, 30$ та виконано відновлення за описаними вище методами [6].

У табл. 3.15 наведено статистичні характеристики, що показують якість побудованих моделей у цілому для випадку № 1 при відновленні значень індексу F10.7, розраховані за формулами (3.1), (3.2), (3.3) на проміжку $N = 1, \dots, 44$ [6]. У цій таблиці моделі позначено таким чином: експоненційне згладжування – ЕЗ та адаптивне ковзне двобічне експоненційне згладжування –

АКДЕЗ. У табл. 3.16 наведено статистичні характеристики моделей аналогічно до табл. 3.15 для випадку № 2 при відновленні значень індексу F10.7 [6].

Таблиця 3.15

Статистичні характеристики моделей

Статистична характеристика	ЕЗ	АКДЕЗ
R^2	0,9162	0,9241
SSE	95,3291	91,7328
DW	1,8552	1,9171

Таблиця 3.16

Статистичні характеристики моделей

Статистична характеристика	ЕЗ	АКДЕЗ
R^2	0,9271	0,9423
SSE	94,5172	88,1214
DW	1,8632	1,9364

За значеннями статистичних характеристик, наведених у табл. 3.15, можна зробити такі висновки: значення коефіцієнта детермінації R^2 для адаптивного ковзного двобічного експоненційного згладжування є ближчим до ідеального значення 1 та більшим на 0,855%, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для розробленого методу є меншою на 3,773%, ніж для експоненційного згладжування; статистика Дарбіна-Уотсона DW для адаптивного ковзного двобічного експоненційного згладжування є ближчою до ідеального значення 2 та більшою на 3,229%, ніж для експоненційного згладжування.

За значеннями статистичних характеристик, наведених у табл. 3.16, можна зробити такі висновки: значення коефіцієнта детермінації R^2 для адаптивного ковзного двобічного експоненційного згладжування є ближчим до ідеального значення 1 та більшим на 1,613%, ніж для експоненційного згладжування; сума квадратів похибок моделі SSE для розробленого методу є меншою на 6,767%, ніж для експоненційного згладжування; статистика Дарбіна-Уотсона DW для адаптивного ковзного двобічного експоненційного згладжування є ближчою до ідеального значення 2 та більшою на 3,78%, ніж для експоненційного згладжування.

За значеннями статистичних характеристик, наведених у табл. 3.15 та табл. 3.16, можна зробити висновок, що метод адаптивного ковзного двобічного експоненційного згладжування показує перевагу перед методом експоненційного згладжування при відновленні пропущених значень індексу F10.7. Тому можна стверджувати, що запропонований метод є ефективним для відновлення пропущених значень часових рядів за різних початкових умов.

3.4.4. Приклад прогнозування значень часових рядів з використанням методу адаптивного ковзного двобічного експоненційного згладжування

У роботі [6] на основі вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU з січня 2013 року до квітня 2016 року (41 значення) [70] виконується прогнозування значень на 3 наступних місяці. Прогнозування у роботі [6] виконується за допомогою методу адаптивного ковзного двобічного експоненційного згладжування за формулами (3.14), (3.15), (3.2), (3.4) та методу, що наведений у роботі [28] та описується формулами (1.16) – (1.18).

Обчислюються значення прогнозних характеристик для побудованих моделей: середньоквадратичної похибки $RMSE$, середньої відсоткової абсолютної похибки $MAPE$ та коефіцієнта Тейла U за формулами (2.22),

(2.23), (2.24) відповідно. У табл. 3.17 наведено прогнозні характеристики, розраховані на проміжку $N=42, \dots, 44$ [6]. У цій таблиці моделі позначено таким чином: експоненційний підхід, що описується формулами (1.16)–(1.18) – ЕП та адаптивне ковзне двобічне експоненційне згладжування – АКДЕЗ.

Таблиця 3.17

Значення прогнозних характеристик для моделей

Прогнозна характеристика	ЕП	АКДЕЗ
RMSE	8,0462	6,1278
MAPE	8,2261	6,5432
U	0,0047	0,0045

За значеннями прогнозних характеристик у табл. 3.17 можна зробити висновок, що метод адаптивного ковзного двобічного експоненційного згладжування показує перевагу перед традиційним експоненційним підходом при прогнозуванні: значення середньоквадратичної похибки RMSE є меншим для розробленого методу на 23,842%, ніж для експоненційного згладжування; значення середньої відсоткової абсолютної похибки MAPE є меншим для методу адаптивного ковзного двобічного експоненційного згладжування на 20,458%, ніж для експоненційного згладжування; значення коефіцієнта Тейла U є меншим для розробленого методу на 4,255%, ніж для експоненційного згладжування.

3.5. Висновки до розділу 3

Розроблено метод двобічного експоненційного згладжування, що складається з процедури традиційного експоненційного згладжування та застосування процедури експоненційного згладжування до отриманих оцінок у зворотному часі. Додавання другого етапу для додаткової фільтрації похибок у

всіх попередніх точках у методі двобічного експоненційного згладжування дає можливість отримати більш точні оцінки значень ряду. Також розроблено інтегральний критерій адекватності моделі, запропоновано критерій вибору оптимального значення параметра згладжування. Виконано імітаційне моделювання для перевірки ефективності методу двобічного експоненційного згладжування. Розраховуються такі статистичні характеристики для оцінювання якості побудованих моделей у цілому: коефіцієнт детермінації R^2 , сума квадратів похибок моделі SSE , статистика Дарбіна-Уотсона DW . Отримані результати моделювання показують, що двобічне експоненційне згладжування дозволяє отримати ближчі до вимірів і менш варіабельні оцінки значень ряду та показує перевагу перед експоненційним згладжуванням за всіма статистичними характеристиками.

Описано застосування методу двобічного експоненційного згладжування для відновлення істинних закономірностей та прогнозування часових рядів. Розроблено алгоритм для відновлення пропущених значень часових рядів з використанням методу двобічного експоненційного згладжування. У цьому розробленому алгоритмі передбачено можливості відновлення пропущених значень часових рядів за наявності різних початкових умов. Виконано відновлення пропущених значень середньодобових цін на цинк за даними Лондонської біржі металів за допомогою таких методів [3]: розробленого методу двобічного експоненційного згладжування [3]; заміни пропущених значень середніми арифметичними значеннями [23]; відновлення за допомогою методу експоненційного згладжування з використанням в якості відомого значення виміру останнього значення перед пропущеним значенням та з початковим наближенням, рівним середньому арифметичному [23] серед усіх відомих значень; за допомогою авторегресійного підходу, запропонованого у роботі [24]. За значеннями наведених вище статистичних характеристик зроблено висновок, що розроблений метод двобічного експоненційного згладжування показує перевагу перед усіма іншими методами. При відновленні

пропущених значень часових рядів за різних початкових умов показано ефективність застосування методу двобічного експоненційного згладжування.

Розроблено метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Створений метод застосовано для різних випадків відновлення штучних пропусків та фактично пропущених значень показників сталого розвитку. У всіх випадках розроблений метод показує перевагу порівняно із заміною пропущених значень за допомогою середніх арифметичних та експоненційного згладжування.

Також розроблено метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. З використанням створеного методу та традиційного експоненційного згладжування виконується прогнозування значень ряду індексу середньої кількості дітей, що навчаються у школі. Після побудови математичних моделей обчислено значення прогнозних характеристик, а саме, значення середньоквадратичної похибки RMSE, середньої відсоткової абсолютної похибки MAPE та коефіцієнта Тейла U. Розроблений метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування показує перевагу перед методом експоненційного згладжування за значеннями усіх прогнозних характеристик.

Розроблено метод ковзного двобічного експоненційного згладжування. У роботі з використанням методу ковзного двобічного експоненційного згладжування розроблено алгоритм для відновлення пропущених значень ряду. Виконано відновлення пропущених значень з використанням створеного методу та методу експоненційного згладжування. За значеннями статистичних характеристик розроблений метод показує перевагу перед методом експоненційного згладжування.

Розроблено критерій близькості для оцінювання моделей під час відновлення істинних закономірностей розвитку часових рядів. Виконано відновлення істинних закономірностей розвитку індексу F10.7 з застосуванням

методу ковзного двобічного експоненційного згладжування та 13-місячного ковзного середнього. Застосовуючи критерій близькості, зроблено висновок, що метод ковзного двобічного експоненційного згладжування показує перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей індексу F10.7.

Метод ковзного двобічного експоненційного згладжування використано для прогнозування індексу F10.7 за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США та виконано порівняння з експоненційним підходом, запропонованим у роботі [28]. За значеннями прогнозних характеристик метод ковзного двобічного експоненційного згладжування показує перевагу порівняно із традиційним експоненційним підходом при прогнозуванні.

Розроблено метод адаптивного ковзного двобічного експоненційного згладжування. Його застосовано для відновлення істинних закономірностей ряду чисел Вольфа та індексу F10.7, виконано порівняння з методом 13-місячного ковзного середнього. За критерієм близькості зроблено висновок, що метод адаптивного ковзного двобічного експоненційного згладжування показує перевагу порівняно з методом 13-місячного ковзного середнього при відновленні істинних закономірностей чисел Вольфа та індексу F10.7.

Розроблено алгоритм відновлення пропущених значень з використанням методу адаптивного ковзного двобічного експоненційного згладжування. Виконано відновлення пропущених значень індексу F10.7 з використанням створеного алгоритму та методу експоненційного згладжування. За значеннями статистичних характеристик зроблено висновок, що алгоритм відновлення пропущених значень з використанням методу адаптивного ковзного двобічного експоненційного згладжування показує перевагу порівняно з методом експоненційного згладжування при відновленні пропущених значень індексу F10.7.

Метод адаптивного ковзного двобічного експоненційного згладжування застосовано також для прогнозування значень сонячних радіофлюксів при

довжині хвилі 10,7 см та виконано порівняння з експоненційним підходом, запропонованим у роботі [28]. За значеннями прогнозних характеристик метод адаптивного ковзного двобічного експоненційного згладжування показує перевагу порівняно із традиційним експоненційним підходом при прогнозуванні.

РОЗДІЛ 4

РОЗРОБКА СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОРГАНІЗАЦІЇ, ЩО ЗАЙМАЄТЬСЯ АНАЛІЗОМ ТА ПРОГНОЗУВАННЯМ ЧАСОВИХ РЯДІВ

Аналіз та прогнозування великої кількості даних у формі часових рядів нерозривно пов'язані з розробкою інформаційних систем для їх автоматизованої обробки. Тому системні задачі створення та вдосконалення існуючих систем підтримки прийняття рішень для аналізу та прогнозування часових рядів є актуальними на сьогоднішній день.

Дослідженням питання розробки систем підтримки прийняття рішень займалось багато науковців. Зокрема, детальний алгоритм побудови системи підтримки прийняття рішень для аналізу та прогнозування часових рядів наведено у роботі [43]. Також у роботі [43] наведено схему процесу прогнозування часових рядів. У роботі [44] виконано дослідження стосовно розробки системи підтримки прийняття рішень для прогнозування динаміки стаціонарних та нестаціонарних процесів різної природи. Результати наукових досліджень та практичних розробок методів експертного оцінювання та методів побудови систем підтримки прийняття рішень наведено у роботі [46]. Приклади типових систем підтримки прийняття рішень, архітектуру систем підтримки прийняття рішень та баз даних, що використовуються у системах підтримки прийняття рішень, наведено у роботі [47].

Таким чином, багатьма науковцями виконано дослідження стосовно системної задачі розробки архітектури систем підтримки прийняття рішень для аналізу та прогнозування часових рядів. Проте подальше вдосконалення систем підтримки прийняття рішень та розробка нових таких систем з реалізацією нових методів для аналізу та прогнозування часових рядів є актуальними, оскільки це дасть можливість отримувати більш точні оцінки значень та виконувати більш швидку обробку часових рядів.

4.1. Застосування системного підходу до розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів

Розробка системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів, є складною системною задачею. Для розробки такої системи повинен бути виконаний попередній інформаційний аналіз.

Основні цілі інформаційного аналізу зводяться до забезпечення необхідного і технологічно можливого рівня інформаційного забезпечення достовірності та обґрунтованості розв'язання прикладних системних задач [12]. Задачі інформаційного аналізу полягають у створенні методологічного і математичного інструментарію для досягнення поставлених цілей [12]. Саме тому спочатку необхідно визначити архітектуру системи підтримки прийняття рішень.

Важливим є дотримання основних принципів формування системної методології та її практичної реалізації [12]: системної погодженості, процедурної повноти, функціональної ортогональності, інформаційної взаємозалежності, цілеспрямованої відповідності, функціональної раціональності, багатоцільової загальності, багатofакторної адаптивності, процедурної відкритості, раціональної доповнюваності. Ці принципи також повинні бути врахованими під час проектування та розробки СППР.

Для організації, що займається аналізом та прогнозуванням часових рядів, робота за допомогою системи підтримки прийняття рішень повинна забезпечувати можливості обробки даних про співробітників, а також обробку, аналіз та прогнозування часових рядів, тобто становити цілісну системну методологію.

Основними функціями при роботі з розробленою СППР є введення даних про нових співробітників, введення даних про зареєстрованих співробітників – прийняття на посаду або звільнення з посади, введення даних про нові часові

ряди, введення даних про зареєстровані часові ряди, завантаження часових рядів з бази даних або текстового файлу, виконання початкового аналізу часових рядів, відновлення пропущених значень часових рядів, відновлення істинних закономірностей часових рядів, прогнозування часових рядів, побудова графіків отриманих результатів.

В організації, що займається аналізом та прогнозуванням часових рядів, виділяються відділи інформаційних технологій, аналітики, кадрів. Виділяються групи користувачів та надається можливість працювати з СППР працівникам кожного з перерахованих відділів відповідно до основних функцій, що виконуються при роботі з СППР. Таким чином забезпечується принцип інформаційної взаємозалежності, тобто існує погодженість між усіма процедурами системної методології.

Співробітники відділу інформаційних технологій мають право працювати з усіма даними та СППР, щоб забезпечувати повне та правильне функціонування системи. Співробітники відділу аналітики мають право вводити дані про нові та зареєстровані часові ряди, завантажувати часові ряди з бази даних та текстового файлу, виконувати початковий аналіз часових рядів, відновлювати пропущені значення та істинні закономірності розвитку часових рядів, виконувати прогнозування часових рядів, будувати графіки отриманих результатів. Співробітники відділу кадрів мають право вводити дані про нового співробітника, вводити дані про прийняття на посаду зареєстрованого співробітника та вводити дані про звільнення з посади наявного співробітника.

Аналіз та прогнозування часових рядів виконуються відповідно до розробленої системної ієрархії. Структурну схему частини системи підтримки прийняття рішень, що стосується аналізу та прогнозування часових рядів, наведено на рис. 4.1.

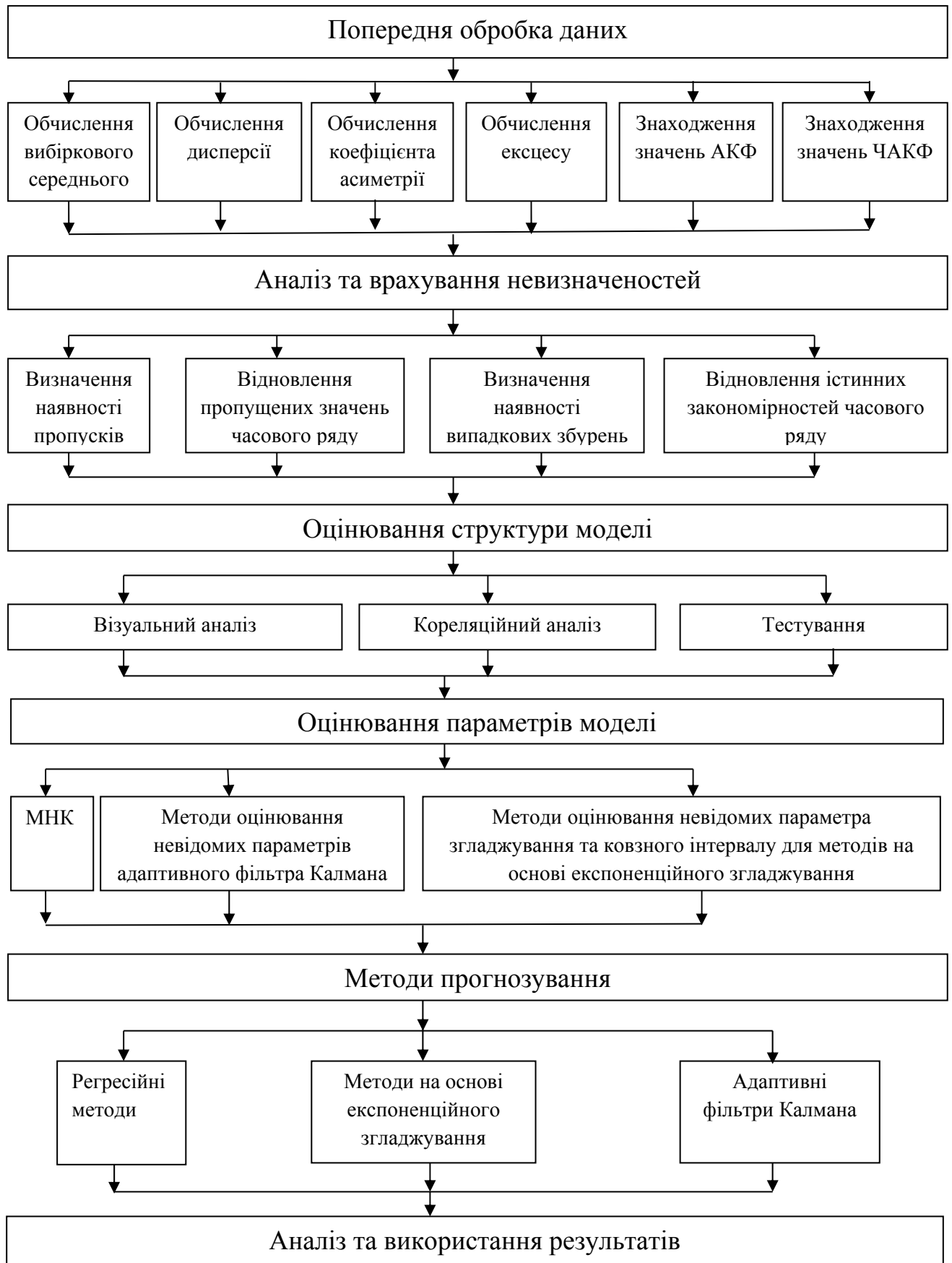


Рис. 4.1. Структурна схема частини аналізу та прогнозування часових рядів у системі підтримки прийняття рішень

Наведена на рис. 4.1 схема повністю описує процес аналізу та прогнозування часових рядів з використанням системного підходу. Також на цій схемі показано, що забезпечується виконання принципу процедурної повноти, тобто системна методологія забезпечує виконання усіх процедур, починаючи від формалізації формулювання системної задачі до верифікації отриманих результатів її розв'язання.

Таким чином, використання системного підходу під час розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів, є важливим, зокрема, виконання інформаційного аналізу перед розробкою СППР.

4.1.1. Бізнес-правила

Бізнес-правила описують основні правила роботи з СППР для співробітників організації, що займається аналізом та прогнозуванням часових рядів.

1. Співробітник кожного з відділів має право працювати з даними у рамках визначених для відділу прав доступу до даних.
2. Одна фізична особа може працювати тільки на одній посаді одночасно.
3. Вводити дані про звільнення фізичної особи з посади можна тільки у тому випадку, якщо фізична особа працює у даний момент на цій посаді.
4. Не можна додавати у базу даних запис про часовий ряд, якщо вже існує запис про нього.
5. Не можна додавати у базу даних запис про фізичну особу, якщо вже існує запис про цю фізичну особу.
6. Не можна додавати у базу даних запис про те, що фізична особа працює на посаді з певним типом зайнятості, якщо вже є запис про це.

Наведені бізнес-правила забезпечують виконання принципів системної погодженості та функціональної раціональності.

4.1.2. Матриці подій для співробітників відділів організації

Роботу відділів аналітики та кадрів показано у вигляді матриць подій, наведених у табл. 4.1 і табл. 4.2.

Таблиця 4.1

Матриця подій для співробітників відділу аналітики

№	Подія	Тип	Дія (реакція)	Перехід
1	Поява даних про часовий ряд.	N	Якщо часовий ряд зареєстрований, то ... Інакше	3 2
2	Реєстрація часового ряду.	N	Внесення нового запису про часовий ряд до бази даних.	
3	Внесення даних про часовий ряд.	N	Внесення нового запису про дані про часовий ряд до бази даних.	
4	Аналіз та обробка даних про часовий ряд.	N	Здійснення аналізу та обробки даних про часовий ряд, якщо ці дані є. Інакше – відмова у виконанні.	

Таблиця 4.2

Матриця подій для співробітників відділу кадрів

№	Подія	Тип	Дія (реакція)	Перехід
1	Дані про прийняття співробітника на посаду.	N	Якщо співробітник зареєстрований, то... Інакше	3 2
2	Реєстрація нового співробітника.	N	Внесення нового запису про фізичну особу у базу даних.	
3	Прийняття співробітника на посаду.	N	Внесення нового запису про фізичну особу на посаді у базу даних.	
4	Звільнення співробітника з посади.	N	Внесення даних про звільнення з посади співробітника до бази даних, якщо такий співробітник працював.	

Відділ інформаційних технологій має повний сумарний набір функцій, які мають відділи аналітики та кадрів, тому для цього відділу матриця подій окремо не будується. Наведені матриці подій показують, що у даній СППР виконуються принципи системної погодженості та процедурної повноти.

4.2. Проектування бази даних

Проектування бази даних є початковим етапом розробки бази даних та системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів.

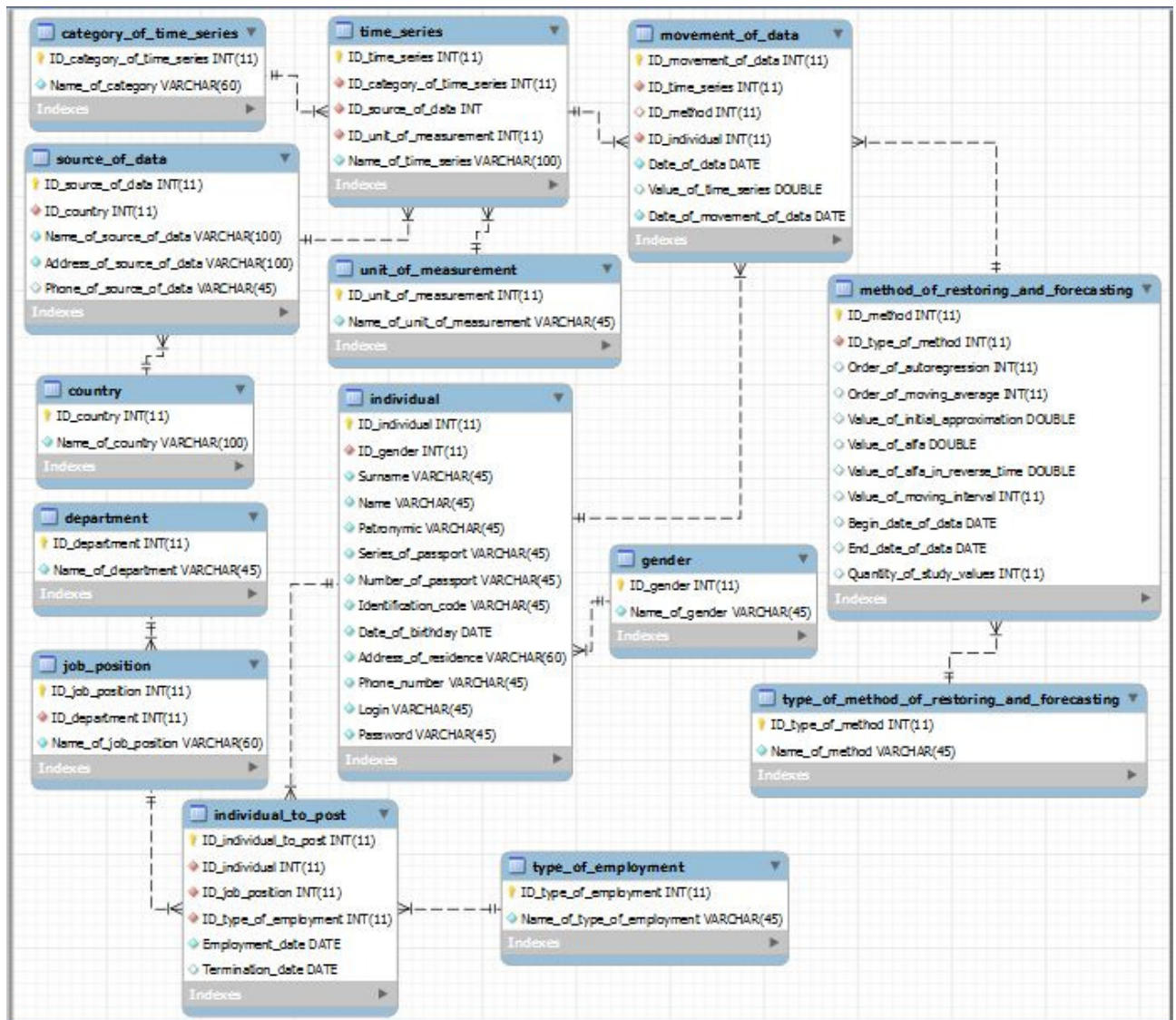


Рис. 4.2. ER-діаграма фізична загальна бази даних

На рис. 4.2 зображено ER-діаграму фізичну загальну бази даних, що є результатом виконаного проектування бази даних. Засобами адміністрування бази даних є програми MySQL Workbench 6.1 CE та MySQL Server 5.5. Ці програми дозволяють легко створювати таблиці, керувати введенням та оновленням даних.

Таким чином, проектування бази даних є важливим етапом для розробки системи підтримки прийняття рішень та після виконання цього етапу може відбуватись безпосередня програмна реалізація СППР.

4.3. Опис програмної реалізації системи підтримки прийняття рішень

Створення архітектури системи підтримки прийняття рішень перед її подальшою програмною реалізацією є важливою системною задачею. У роботі [9] запропоновано архітектуру системи підтримки прийняття рішень, що складається з трьох блоків: попереднього аналізу даних, побудови моделей за даними і прогнозування за моделями. Загальну архітектуру системи підтримки прийняття рішень, що складається з блоків попереднього аналізу даних, відновлення закономірностей розвитку часових рядів, прогнозування часових рядів та графічного відображення отриманих результатів створено у роботі [4].

У даній роботі виконано подальші дослідження та розроблено архітектуру системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів. СППР містить такі закладки: «Авторизація користувача», «Введення даних про нових співробітників», «Введення даних про зареєстрованих співробітників», «Введення даних про нові часові ряди», «Введення даних про зареєстровані часові ряди», «Завантаження часових рядів», «Початковий аналіз часових рядів», «Відновлення пропущених значень часових рядів», «Відновлення істинних закономірностей часових рядів», «Прогнозування часових рядів», «Графіки отриманих результатів». Архітектуру СППР наведено на рис. 4.3. Наведена архітектура СППР показує, що під час розробки СППР було дотримано основних принципів формування системної методології та її практичної реалізації, описаних у роботі [12]: системної погодженості, процедурної повноти, функціональної ортогональності, інформаційної взаємозалежності, цілеспрямованої відповідності, функціональної раціональності, багатоцільової загальності, багатофакторної адаптивності, процедурної відкритості, раціональної доповнюваності. Також досягнуто основні цілі інформаційного аналізу, а саме, забезпечено необхідний і технологічно можливий рівень інформаційного забезпечення достовірності та обґрунтованості розв'язання прикладних системних задач. Також виконано

поставлені задачі інформаційного аналізу, тобто створено методологічний і математичний інструментарій для досягнення поставлених цілей під час розробки СППР як цілісної системної методології.

Програмну реалізацію СППР виконано з використанням мови програмування Java 8 та основні графічні структурні компоненти виконано з використанням бібліотеки SWING [88, 89].

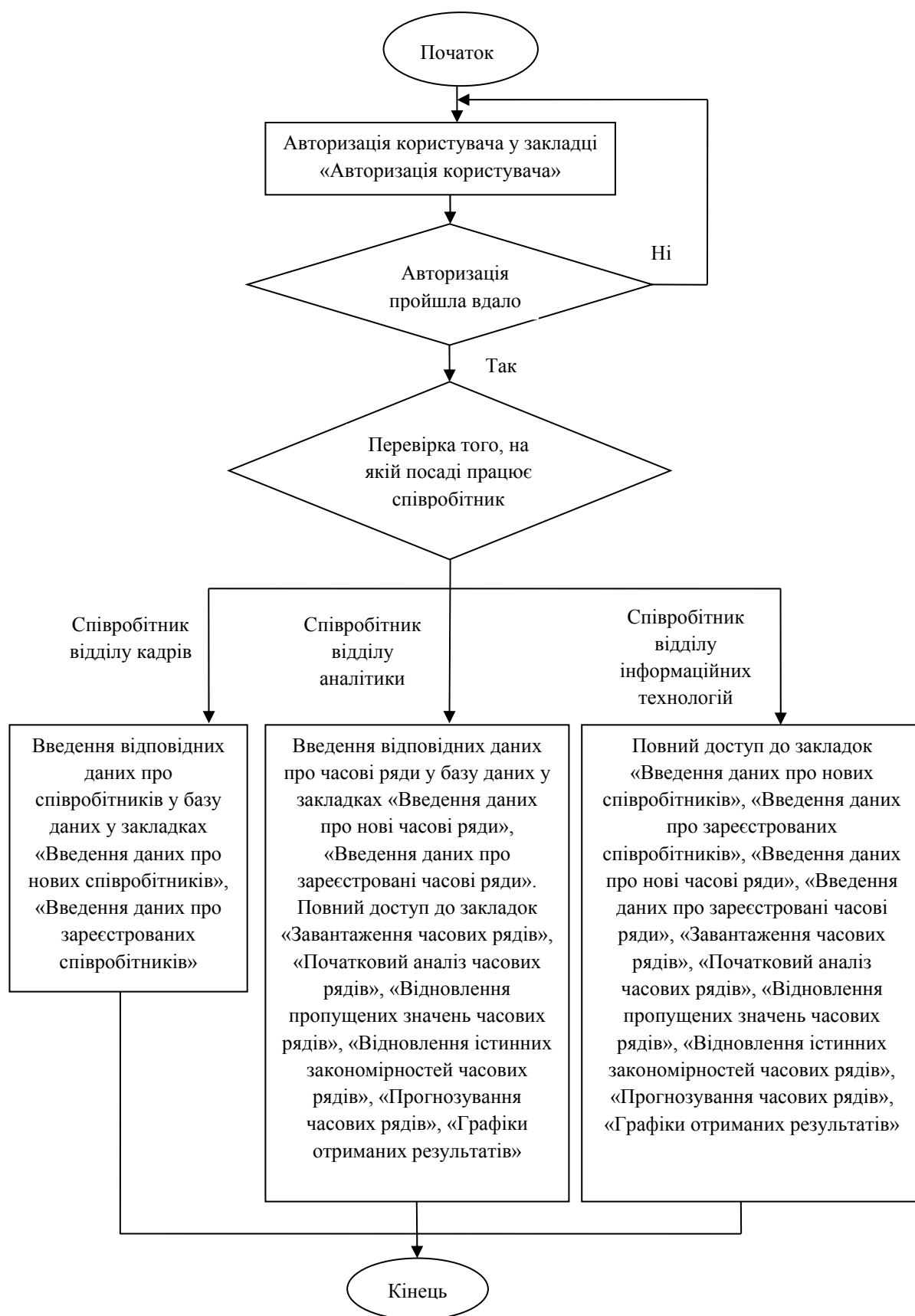


Рис. 4.3. Архітектура системи підтримки прийняття рішень

Для того, щоб розпочати роботу у системі підтримки прийняття рішень, користувачу необхідно пройти процедуру авторизації. Для цього необхідно у відповідних полях закладки «Авторизація користувача», зображеної на рис. 4.4, ввести логін, пароль та натиснути на кнопку «Авторизація».

Рис. 4.4. Закладка «Авторизація користувача»

Якщо авторизація пройшла вдало, то відповідно до посади, на якій працює співробітник, стануть доступними інші закладки системи підтримки прийняття рішень. Якщо авторизація пройшла невдало, то з'являється повідомлення «Неправильно введено логін та пароль.», після чого користувач може повторити спробу авторизації. Кнопка «Очистити» закладки «Авторизація користувача» призначена для очищення полів з написами «Логін» та «Пароль».

До того часу, поки користувач не зайшов у систему, видимою та активною є тільки закладка «Авторизація користувача». Під час процедури авторизації виконується перевірка того, на якій посаді або посадах працює користувач, тобто застосовується системний принцип інформаційної взаємозалежності. Якщо користувач є співробітником відділу інформаційних технологій, то після авторизації у створеній СППР стають видимими та

доступними всі закладки системи підтримки прийняття рішень, що показано на рис. 4.5. Видимість усіх закладок є необхідною умовою для можливості виконання у повному обсязі роботи співробітниками відділу інформаційних технологій.

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів Відновлення істинних закономірностей часових рядів Прогнозування часових рядів Графіки отриманих результатів

Введення даних про нові часові ряди Введення даних про зареєстровані часові ряди Завантаження часових рядів Початковий аналіз часових рядів

Авторизація користувача Введення даних про нових співробітників Введення даних про зареєстрованих співробітників

Введення даних про нових співробітників

Назва статі: Male

Прізвище: Tkachenko

Ім'я: Yurii

По-батькові: Pavlovych

Серія паспорта: ME

Номер паспорта: 655421

Ідентифікаційний код: 3000212824

Дата народження: 08.04.1987

Адреса місця проживання: Kyiv, str. Shuliavska 15, apt 11

Номер телефону: 0673425421

Логін: yurii_tkachenko

Пароль:

Підтвердити пароль:

Підтвердити Очистити

Рис. 4.5. Вигляд закладки «Введення даних про нових співробітників» та видимість інших закладок, коли авторизований співробітник відділу інформаційних технологій

Якщо користувач є співробітником відділу аналітики, то після авторизації у створеній СППР стають видимими та доступними такі закладки: «Авторизація користувача», «Введення даних про нові часові ряди», «Введення даних про зареєстровані часові ряди», «Завантаження часових рядів», «Початковий аналіз часових рядів», «Відновлення пропущених значень часових рядів», «Відновлення істинних закономірностей часових рядів», «Прогнозування часових рядів», «Графіки отриманих результатів». Видимість закладок для співробітників відділу аналітики показано на рис. 4.6. Видимість

наведених вище закладок забезпечує можливість роботи у повному обсязі для співробітників відділу аналітики.

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів | Відновлення істинних закономірностей часових рядів | Прогнозування часових рядів | Графіки отриманих результатів

Авторизація користувача | **Введення даних про нові часові ряди** | Введення даних про зареєстровані часові ряди | Завантаження часових рядів | Початковий аналіз часових рядів

Введення даних про нові часові ряди

Назва країни: United Kingdom of Great Britain and Northern Ireland

Назва джерела даних: London metal exchange

Адреса джерела даних: https://www.lme.com

Телефон джерела даних: +44(0)2071138888

Назва категорій часового ряду: Financial data

Назва часового ряду: Zinc daily average price

Одиниця виміру: USD per tonna

Підтвердити | Очистити

Рис. 4.6. Вигляд закладки «Введення даних про нові часові ряди» та видимість інших закладок, коли авторизований співробітник відділу аналітики

Якщо користувач є співробітником відділу кадрів, то видимими та доступними після успішної авторизації у розробленій СППР будуть закладки «Авторизація користувача», «Введення даних про нових співробітників», «Введення даних про зареєстрованих співробітників», що показано на рис. 4.7. Видимість наведених закладок забезпечує можливість виконання роботи співробітниками відділу кадрів.

Рис. 4.7. Вигляд закладки «Введення даних про нових співробітників» та видимість інших закладок, коли авторизований співробітник відділу кадрів

На рис. 4.7 показано закладку «Введення даних про нових співробітників», на якій реалізовано можливість введення даних про нових співробітників та запис введених даних до бази даних після натискання на кнопку «Підтвердити». Співробітник повинен вибрати назву статі зі значень у списку та дату народження з використанням стандартного компоненту, всі інші значення повинні бути введені вручну. Показується повідомлення про помилку у випадку, якщо введено дані не у всі поля або у деякі поля закладки введено неправильні дані. Після натискання на кнопку «Очистити» відбувається очищення усіх заповнених полів на цій закладці.

На рис. 4.8 показано закладку «Введення даних про зареєстрованих співробітників», на якій реалізовано можливість прийняття на посаду вже зареєстрованого співробітника, який ще не працює на жодній з посад, після натискання на кнопку «Прийняти на посаду». На рис. 4.9 показано, як на цій закладці можна ввести інформацію про звільнення з посади зареєстрованого співробітника після натискання на кнопку «Звільнити з посади», якщо цей

співробітник працює та зареєстрований на посаді. Після натискання на кнопку «Очистити» усі заповнені поля на цій закладці очищуються. Показується відповідне повідомлення про помилку у випадку, якщо під час прийняття на посаду або звільнення з посади співробітника введено дані не у всі необхідні поля.

Кандидатська дисертація Братусь Олена

Авторизація користувача Введення даних про нових співробітників Введення даних про зареєстрованих співробітників

Введення даних про зареєстрованих співробітників

Назва статі: Male

Прізвище: Fedorchenko

Серія паспорта: ME

Номер паспорта: 423487

Назва відділу: Administrators

Назва посади: Administrator

Назва типу зайнятості: Full-time employment

Дата прийому на посаду: 11.05.2018

Дата звільнення з посади: 11.05.2018

Прийняти на посаду Звільнити з посади Очистити

Рис. 4.8. Вигляд закладки «Введення даних про зареєстрованих співробітників», коли співробітника приймають на посаду

Кандидатська дисертація Братусь Олена

Авторизація користувача Введення даних про нових співробітників Введення даних про зареєстрованих співробітників

Введення даних про зареєстрованих співробітників

Назва статі: female

Прізвище: Tkachenko

Серія паспорта: MA

Номер паспорта: 702542

Назва відділу: Human resources

Назва посади:

Назва типу зайнятості: full-time employment

Дата прийому на посаду: 11.05.2018

Дата звільнення з посади: 11.05.2018

Прийняти на посаду Звільнити з посади Очистити

Рис. 4.9. Вигляд закладки «Введення даних про зареєстрованих співробітників», коли співробітника звільняють з посади

На рис. 4.6 наведено вигляд закладки «Введення даних про нові часові ряди», на якій можна ввести дані про новий часовий ряд та підтвердити це після натискання на кнопку «Підтвердити». На рис. 4.6 показано приклад заповнення форми для випадку, коли вводяться дані про середньодобові ціни Лондонської біржі металів на цинк у американських доларах за тонну у робочі дні [52]. Під час введення даних про новий часовий ряд співробітник може обирати наявні значення зі списків вибору у тих полях, де це можливо, або вводити нові значення. Після натискання на кнопку «Очистити» усі поля форми очищуються від заповнених значень. Показується відповідне повідомлення про помилку у випадку, якщо введено дані не у всі поля або у деякі поля закладки введено неправильні дані.

На рис. 4.10 наведено вигляд закладки «Введення даних про зареєстровані часові ряди», на якій можна ввести дані про зареєстрований часовий ряд та підтвердити це після натискання на кнопку «Підтвердити». На рис. 4.6 показано приклад заповнення форми для випадку, коли вводяться дані

про середньодобову ціну Лондонської біржі металів на цинк у американських доларах за тонну за 04.01.2017 [52]. Вручну вводиться тільки значення часового ряду, всі інші поля обираються зі списків можливих значень, а також дата обирається за допомогою використання компонента календаря. Після натискання на кнопку «Очистити» усі поля форми очищуються від заповнених значень. Показується відповідне повідомлення про помилку у випадку, якщо введено дані не у всі поля або у деякі поля закладки введено неправильні дані.

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів | Відновлення істинних закономірностей часових рядів | Прогнозування часових рядів | Графіки отриманих результатів | Авторизація користувача | Введення даних про нові часові ряди | **Введення даних про зареєстровані часові ряди** | Завантаження часових рядів | Початковий аналіз часових рядів

Введення даних про зареєстровані часові ряди

Назва країни: United Kingdom of Great Britain and Northern Ireland

Назва джерела даних: London metal exchange

Назва категорії часового ряду: Financial data

Одиниця виміру: USD per tonna

Назва часового ряду: Zinc daily average price

Дата часового ряду: 04.01.2017

Значення часового ряду: 2529.5

Дата введення даних: 21.05.2018

Підтвердити | Очистити

Рис. 4.10. Вигляд закладки «Введення даних про зареєстровані часові ряди»

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів | Відновлення істинних закономірностей часових рядів | Прогнозування часових рядів | Графіки отриманих результатів | Авторизація користувача | Введення даних про нові часові ряди | Введення даних про зареєстровані часові ряди | Завантаження часових рядів | Початковий аналіз часових рядів

Завантаження часового ряду з бази даних або текстового файлу

Назва країни:

Назва джерела даних:

Назва категорій часового ряду:

Одиниця виміру:

Назва часового ряду:

За період з: по:

Назва текстового файлу:

Значення ряду

Дата	Значення ряду
03.01.2017	2552.25
04.01.2017	2529.5
05.01.2017	2608.25
06.01.2017	2578.25
09.01.2017	2610.75

Кількість відсутніх у ряді значень:

Поділ вибірки на навчальну: та перевіірочну:

Розмір навчальної вибірки: та перевіірочної:

Рис. 4.11. Вигляд закладки «Завантаження часових рядів», коли часовий ряд завантажено з бази даних

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів | Відновлення істинних закономірностей часових рядів | Прогнозування часових рядів | Графіки отриманих результатів | Авторизація користувача | Введення даних про нові часові ряди | Введення даних про зареєстровані часові ряди | Завантаження часових рядів | Початковий аналіз часових рядів

Завантаження часового ряду з бази даних або текстового файлу

Назва країни:

Назва джерела даних:

Назва категорій часового ряду:

Одиниця виміру:

Назва часового ряду:

За період з: по:

Назва текстового файлу:

Значення ряду

Дата	Значення ряду
03.01.2017	2552.25
04.01.2017	2529.5
05.01.2017	
06.01.2017	
09.01.2017	2610.75

Кількість відсутніх у ряді значень:

Поділ вибірки на навчальну: та перевіірочну:

Розмір навчальної вибірки: та перевіірочної:

Рис. 4.12. Вигляд закладки «Завантаження часових рядів», коли часовий ряд завантажено з текстового файлу

На рис. 4.11 зображено закладку «Завантаження часових рядів» для випадку, коли часовий ряд завантажено з бази даних: перед цим потрібно обрати всі дані з запропонованих списків значень, потім обрати часовий період, за який потрібно завантажити дані, та натиснути на кнопку «Завантажити ряд з бази даних». На рис. 4.12 зображено закладку «Завантаження часових рядів» для випадку, коли часовий ряд завантажено з текстового файлу: перед цим потрібно обрати текстовий файл, з якого буде завантажуватись часовий ряд, та натиснути на кнопку «Завантажити ряд з текстового файлу». Після завантаження ряду у необхідному форматі з текстового файлу будуть заповненими і недоступними для редагування всі поля, в яких показується інформація про часовий ряд. На рис. 4.11 та рис. 4.12 показано завантаження даних про середньодобові ціни Лондонської біржі металів на цинк у американських доларах за тонну у робочі дні за період з 03.01.2017 по 15.02.2017 [52], на рис. 4.12 показано інформацію про штучно створені пропуски двох значень цього часового ряду. Можливості завантаження даних з бази даних та з текстового файлу задовольняють системному принципу функціональної ортогональності, тобто є двома незалежними функціями у СППР. Після завантаження часового ряду з бази даних або текстового файлу автоматично заповнюється поле «Кількість відсутніх у ряді значень». Також після завантаження часового ряду користувачеві необхідно виконати поділ вибірки на навчальну та перевірочну за допомогою вибору співвідношення кількості значень у навчальній вибірці зі списку значень «Поділ вибірки на навчальну». Після цього автоматично обчислюється значення співвідношення для перевірочної вибірки та записується у поле для значень перевірочної вибірки, а також записуються числові значення розмірів навчальної та перевірочної вибірок. Після натискання на кнопку «Очистити» усі поля на закладці «Завантаження часових рядів» очищуються.

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів Відновлення істинних закономірностей часових рядів Прогнозування часових рядів Графіки отриманих результатів
 Авторизація користувача Введення даних про нові часові ряди Введення даних про зареєстровані часові ряди Завантаження часових рядів Початковий аналіз часових рядів

Обчислення статистичних характеристик часового ряду

у середнє: 2744.1552
 Дисперсія: 8853.0822
 S: 0.0103
 K: 2.7444

Автокореляційна функція ряду

l	r(l)
0	0.8514
1	0.5836
2	0.4534
3	0.3164
4	0.2113

Часткова автокореляційна функція ряду

l	Ф(l)
0	1.0
1	0.5836
2	0.1711
3	0.2348
4	0.0802

Автокореляційна функція залишків моделі

l	r(l)
0	0.9884
1	0.2319
2	-0.0487
3	-0.1964
4	-0.2711

Часткова автокореляційна функція залишків моделі

l	Ф(l)
0	1.0
1	0.2319
2	-0.1083
3	-0.1722
4	-0.2393

Визначення рекомендованих порядків авторегресії та ковзного середнього

Поріг значень ЧАКФ ряду: 0.1 Рекомендований порядок авторегресії: 3
 Поріг значень ЧАКФ залишків моделі: 0.1 Рекомендований порядок ковзного середнього: 1

Знайти статистичні характеристики ряду Знайти рекомендований порядок АР Знайти рекомендований порядок КС Очистити

Рис. 4.13. Вигляд закладки «Початковий аналіз часових рядів»

На рис. 4.13 показано вигляд закладки «Початковий аналіз часових рядів», показано обчислені значення для початкового аналізу даних про середньодобові ціни Лондонської біржі металів на цинк у американських доларах за тонну у робочі дні за період з 03.01.2017 по 15.02.2017 [52]. На цій закладці користувач має можливість обчислити статистичні характеристики часового ряду, а саме, знайти значення вибіркового середнього, дисперсії, коефіцієнта асиметрії, ексцесу. Вибіркове середнє обчислюється за формулою (2.12). Дисперсія обчислюється так [23]:

$$\sigma^2 = \frac{1}{N-1} \times \sum_{k=1}^N [y(k) - \bar{y}]^2, \quad (4.1)$$

де σ^2 – дисперсія; N – кількість значень ряду; $y(k)$ – значення ряду в k -й момент часу; \bar{y} – вибіркове середнє.

Коефіцієнт асиметрії розраховується таким чином [23]:

$$S = \frac{\sum_{k=1}^N [y(k) - \bar{y}]^3}{N \times \sigma^3}, \quad (4.2)$$

де σ – середньоквадратичне відхилення.

Ексцес обчислюється так [23]:

$$K = \frac{\sum_{k=1}^N [y(k) - \bar{y}]^4}{N \times \sigma^4}. \quad (4.3)$$

Також на закладці «Початковий аналіз часових рядів» користувач має можливість обчислити значення автокореляційної та часткової автокореляційної функцій ряду. Значення усіх перерахованих вище статистичних характеристик, а також АКФ та ЧАКФ ряду розраховуються після натискання на кнопку «Знайти статистичні характеристики ряду». Значення автокореляційної функції ряду розраховуються таким чином [24]:

$$r(l) = \frac{\frac{1}{N-1} \sum_{k=l+1}^N \{[y(k) - \bar{y}] \times [y(k-l) - \bar{y}]\}}{\sigma_y^2}, \quad l = 0, 1, 2, \dots, M, \quad \text{де } M \leq \frac{N}{3}, \quad (4.4)$$

де $r(l)$ – l -те значення автокореляційної функції ряду; σ_y^2 – дисперсія.

Значення часткової автокореляційної функції ряду розраховуються так [24]:

$$\Phi_k = r(1), \quad \text{якщо } k = 1, \quad (4.5)$$

де Φ_k – значення часткової автокореляційної функції ряду; $r(l)$ – значення автокореляційної функції ряду;

$$\Phi_k = \frac{r(k) - \sum_{j=1}^{k-1} \Phi_{k-1,j} \times r(k-j)}{1 - \sum_{j=1}^{k-1} \Phi_{k-1,j} \times r(j)}, \quad \text{якщо } k > 1, \quad (4.6)$$

де Φ_k , $\Phi_{k-1,j}$ – відповідні значення часткової автокореляційної функції ряду; $r(k)$, $r(k-j)$, $r(j)$ – відповідні значення автокореляційної функції ряду.

У створеній СППР користувач має можливість обрати рекомендований поріг значень ЧАКФ для ряду та розрахувати рекомендований порядок авторегресії після натискання на кнопку «Знайти рекомендований порядок АР». Також користувач має можливість розрахувати значення автокореляційної та

часткової автокореляційної функції залишків моделі, після цього обрати рекомендований поріг значень ЧАКФ залишків моделі та визначити рекомендований порядок ковзного середнього після натискання на кнопку «Знайти рекомендований порядок КС». Можливості автоматизованого вибору порядку авторегресії та ковзного середнього дозволяють користувачу виконувати швидкий автоматизований аналіз вхідних даних за допомогою СППР. Значення АКФ та ЧАКФ залишків моделі розраховуються аналогічно до цих функцій для значень ряду за формулами (4.4), (4.5), (4.6). Після натискання на кнопку «Очистити» відбувається очищення усіх полів закладки «Початковий аналіз часових рядів».

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів | Відновлення істинних закономірностей часових рядів | Прогнозування часових рядів | Графіки отриманих результатів

Авторизація користувача | Введення даних про нові часові ряди | Введення даних про зареєстровані часові ряди | Завантаження часових рядів | Початковий аналіз часових рядів

Середнє арифметичне

Дата	Значення ряду
01.01.2013	127.1
01.02.2013	104.4
01.03.2013	111.2
01.04.2013	110.7

Відновити пропуски | Записати у БД | Очистити

Експоненційне згладжування

Параметр згладжування: 0.9

Дата	Значення ряду
01.01.2013	127.1
01.02.2013	104.4
01.03.2013	111.2
01.04.2013	110.52

Відновити пропуски | Записати у БД | Очистити

Двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі: 0.9

у зворотному часі: 0.1

Дата	Значення ряду
01.01.2013	127.1
01.02.2013	104.4
01.03.2013	111.2
01.04.2013	125.785

Відновити пропуски | Записати у БД | Очистити

Ковзне двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі: 0.9

у зворотному часі: 0.1

Дата	Значення ряду
01.01.2013	127.1
01.02.2013	104.4
01.03.2013	111.2
01.04.2013	111.2

Відновити пропуски | Записати у БД | Очистити

Адаптивне ковзне двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі: 0.9

у зворотному часі: 0.1

Ковзний період: 11

Дата	Значення ряду
01.01.2013	127.1
01.02.2013	104.4
01.03.2013	111.2
01.04.2013	111.2

Відновити пропуски | Записати у БД | Очистити

Рис. 4.14. Вигляд закладки «Відновлення пропущених значень часових рядів»

На рис. 4.14 показано вигляд закладки «Відновлення пропущених значень часових рядів». На рис. 4.14 показано відновлення двох штучно пропущених значень за квітень та травень 2013 року на основі вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU з січня 2013 року до березня 2016 року за даними офіційного сайту

Національного управління океанічних і атмосферних досліджень США [70]. На закладці «Відновлення пропущених значень часових рядів» надається можливість відновити пропущені значення часового ряду за допомогою заміни пропущених значень середніми арифметичними значеннями за формулою (1.12) та відновити за допомогою методу експоненційного згладжування за формулою (1.13) з використанням в якості відомого значення виміру останнього значення перед пропущеним значенням та з початковим наближенням, рівним середньому арифметичному серед усіх відомих значень. Також на закладці «Відновлення пропущених значень часових рядів» користувачеві надається можливість відновити пропущені значення за допомогою таких розроблених методів: двобічного експоненційного згладжування, що описується рівняннями (1.13) та (3.5); ковзного двобічного експоненційного згладжування, що описується рівняннями (3.12) та (3.13); адаптивного ковзного двобічного експоненційного згладжування, що описується рівняннями (3.14) та (3.15). Для знаходження параметра згладжування на кожному з етапів цих трьох методів використовується критерій мінімізації значень, що описуються рівняннями (3.2) та (3.4). Після натискання на відповідну кнопку «Відновити пропуски» для кожного з наведених вище методів відбувається відновлення пропущених значень часового ряду. Після натискання на відповідну кнопку «Записати у БД» відбувається запис отриманих значень у базу даних для кожного з методів. Після натискання на відповідну кнопку «Очистити» відбувається очищення полів для кожного з методів.

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів | Відновлення істинних закономірностей часових рядів | Прогнозування часових рядів | Графіки отриманих результатів

Авторизація користувача | Введення даних про нові часові ряди | Введення даних про зареєстровані часові ряди | Завантаження часових рядів | Початковий аналіз часових рядів

Тринадцятимісячне ковзне середнє

Дата	Значення ряду
01.02.2014	142.3333
01.03.2014	142.3333
01.04.2014	142.3333
01.05.2014	142.3333

Відновити істинну закономірність | Очистити

Значення SSE

Назва методу	Значення SSE
ТКС	5096.2377
ЕЗ	1903.577
ДЕЗ	888.4045
КДЕЗ	899.9461

Кращі методи: ДЕЗ

Знайти значення SSE | Очистити

Експоненційне згладжування

Вибір початкового наближення

☐ перше значення ряду 170.3

☒ у середнє 126.99615384615382

Параметр згладжування 0.9

Дата	Значення ряду
01.02.2014	126.99615384615382
01.03.2014	147.6096
01.04.2014	144.631
01.05.2014	131.4631

Відновити істинну закономірність | Очистити

Ковзне двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі 0.9

у зворотному часі 0.1

Дата	Значення ряду
01.02.2014	135.5197
01.03.2014	131.6774
01.04.2014	129.4482
01.05.2014	127.7353

Відновити істинну закономірність | Очистити

Двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі 0.9

у зворотному часі 0.1

Дата	Значення ряду
01.02.2014	135.7053
01.03.2014	131.8614
01.04.2014	129.6304
01.05.2014	127.9156

Відновити істинну закономірність | Очистити

Адаптивне ковзне двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі 0.9

у зворотному часі 0.1

Ковзний період: 11

Дата	Значення ряду
01.02.2014	135.6125
01.03.2014	131.7694
01.04.2014	129.5393
01.05.2014	127.8254

Відновити істинну закономірність | Очистити

Рис. 4.15. Вигляд закладки «Відновлення істинних закономірностей часових рядів»

На рис. 4.15 показано вигляд закладки «Відновлення істинних закономірностей часових рядів». На рис. 4.15 показано відновлення істинних закономірностей на основі вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU з лютого 2014 року до березня 2016 року за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70]. На закладці «Відновлення істинних закономірностей часових рядів» надається можливість відновити істинні закономірності часового ряду за допомогою таких традиційних методів: методу тринадцятимісячного ковзного середнього, що описується формулою (1.15); методу експоненційного згладжування за формулою (1.13). Також на закладці «Відновлення істинних закономірностей часових рядів» користувачеві надається можливість відновити пропущені значення за допомогою таких розроблених методів: двобічного експоненційного згладжування, що описується рівняннями (1.13) та (3.5); ковзного двобічного експоненційного згладжування, що описується рівняннями

(3.12) та (3.13); адаптивного ковзного двобічного експоненційного згладжування, що описується рівняннями (3.14) та (3.15). Для знаходження параметра згладжування на кожному з етапів цих трьох методів використовується критерій мінімізації значень, що описуються рівняннями (3.2) та (3.4). Відновлення істинної закономірності за відповідним методом відбувається після натискання на кнопку «Відновити істинну закономірність». На закладці «Відновлення істинних закономірностей часових рядів» користувачеві після натискання на кнопку «Знайти значення SSE» надається можливість розрахувати значення суми квадратів похибок моделі SSE для всіх методів, за якими відбувалось відновлення істинних закономірностей часового ряду, та визначити кращі методи за критерієм мінімізації значення SSE , при цьому назви кращих методів записуються у поле біля заголовку «Кращі методи:». Таким чином, на закладці «Відновлення істинних закономірностей часових рядів» реалізовано системний принцип процедурної повноти, тому що користувач має змогу відновити істинні закономірності за допомогою різних методів та одразу визначити кращі методи. Після натискання на кнопку «Очистити» очищуються значення полів для того методу, біля якого знаходиться ця кнопка у блоці графічних елементів. Також після натискання на кнопку «Очистити», що знаходиться у блоці графічних елементів розрахунку значень SSE та визначення кращих методів, будуть очищені відповідні поля зі значеннями SSE та поле з назвами кращих методів. У випадку, якщо під час вибору або введення параметрів для методу відновлення істинних закономірностей часового ряду користувачем допущено помилку, показується відповідне повідомлення після натискання на кнопку «Відновити істинну закономірність».

Кандидатська дисертація Братусь Олена

Відновлення пропущених значень часових рядів Відновлення істинних закономірностей часових рядів Прогнозування часових рядів Графіки отриманих результатів

Авторизація користувача Введення даних про нові часові ряди Введення даних про зареєстровані часові ряди Завантаження часових рядів Початковий аналіз часових рядів

Авторегресія

Порядок АР: 1

Дата	Значення ряду
03.01.2017	2552.25
04.01.2017	2592.2244
05.01.2017	2626.5335
06.01.2017	2655.9802

Прогнозувати на кроків

Оцінити значення ряду Прогнозувати Записати прогноз у БД Очистити

Авторегресія з ковзним середнім

Порядок АР: 1 Порядок КС: 1

Дата	Значення ряду
03.01.2017	2552.25
04.01.2017	2529.5
05.01.2017	2608.25
06.01.2017	2578.25

Прогнозувати на кроків

Оцінити значення ряду Прогнозувати Записати прогноз у БД Очистити

Експоненційне згладжування

Вибір початкового наближення

☒ перше значення ряду 2552.25

☐ у середнє 2761.671875

Параметр згладжування 0.9

Дата	Значення ряду
03.01.2017	2552.25
04.01.2017	2531.775
05.01.2017	2600.6025
06.01.2017	2580.4853

Прогнозувати на кроків

Оцінити значення ряду Прогнозувати Записати прогноз у БД Очистити

Двобічне експоненційне згладжування

Параметри згладжування:

у прямому часі 0.9

у зворотному часі 0.1

Дата	Значення ряду
03.01.2017	2683.234
04.01.2017	2697.7878
05.01.2017	2716.2337
06.01.2017	2729.0816

Прогнозувати на кроків

Оцінити значення ряду Прогнозувати Записати прогноз у БД Очистити

Алгоритм прогнозування на основі фільтра Калмана

Дата	Значення ряду
03.01.2017	2552.25
04.01.2017	2452.3261
05.01.2017	2357.9926
06.01.2017	2268.807

Прогнозувати на кроків

Оцінити значення ряду Прогнозувати Записати прогноз у БД Очистити

Статистичні характеристики

Назва методу	R2	SSE	DW
АР	0.6805	57492.2009	1.1048
АРКС	1.0	0.0	0.0
ЕЗ	0.9813	592.8217	2.0
ДЕЗ	0.397	184407.6602	0.3722

Кращі методи: АРКС

Знайти статистичні характеристики Очистити

Прогнозні характеристики

Назва методу	RMSE	MAPE, %	U
АР	51.0833	1.4793	0.0088
АРКС	41.9033	1.252	0.0072
ЕЗ	41.8813	1.2294	0.0072
ДЕЗ	41.6597	1.2244	0.0071

Кращі методи: ДЕЗ

Знайти прогнозні характеристики Очистити

Рис. 4.16. Вигляд закладки «Прогнозування часових рядів»

На рис. 4.16 показано вигляд закладки «Прогнозування часових рядів». На рис. 4.16 показано виконану побудову моделей, статистичні та прогнозні характеристики, обчислені для даних про середньодобові ціни Лондонської біржі металів на цинк у американських доларах за тонну у робочі дні за період з 03.01.2017 по 15.02.2017 [52]. На закладці «Прогнозування часових рядів» користувачеві надається можливість прогнозувати значення часових рядів за допомогою таких традиційних методів: авторегресії, що описується формулою (1.14); авторегресії з ковзним середнім, що описується формулою (1.20); методу експоненційного згладжування за формулою (1.13). Також на закладці «Прогнозування часових рядів» користувачеві надається можливість прогнозувати значення часового ряду за допомогою таких розроблених методів: двобічного експоненційного згладжування, що описується рівняннями (1.13) та (3.5); алгоритму № 2 прогнозування на основі фільтра Калмана, що описується рівняннями (2.16) – (2.20), (2.25) та повністю описаний у розділі 2.3.4. Після натискання на кнопку «Оцінити значення ряду» відбувається розрахунок оцінок значень ряду для того методу, біля якого знаходиться ця кнопка у блоці

графічних елементів. Після введення кількості кроків, на які потрібно прогнозувати, та натискання на кнопку «Прогнозувати» відбувається прогнозування за відповідним методом. Після натискання на кнопку «Записати прогноз у БД» прогнозні значення, отримані за відповідним методом, записуються у базу даних. Після натискання на кнопку «Очистити» відбувається очищення полів для того методу, для якого цю кнопку було натиснуто. У випадку, якщо під час вибору або введення параметрів для методу прогнозування користувачем допущено помилку, показується відповідне повідомлення після натискання на кнопки «Оцінити значення ряду» та «Прогнозувати».

На закладці «Прогнозування часових рядів» користувачеві надається можливість розрахунку статистичних характеристик моделей, за якими можна оцінити якість побудованих моделей в цілому, а саме, коефіцієнта детермінації R^2 за формулою (3.1), суми квадратів похибок моделі SSE за формулою (3.2), статистики Дарбіна-Уотсона DW за формулою (3.3). Після натискання на кнопку «Знайти статистичні характеристики» відбувається розрахунок цих статистичних характеристик та запис назв кращих методів, визначених за значеннями цих статистичних характеристик, у полі з заголовком «Кращі методи:». Після натискання на відповідну кнопку «Очистити» відбувається очищення значень наведених статистичних характеристик.

Після побудови математичних моделей на закладці «Прогнозування часових рядів» користувачеві надається можливість розрахунку значень статистичних характеристик, які дають можливість визначити якість прогнозу, а саме, значення середньоквадратичної похибки $RMSE$, середньої відсоткової абсолютної похибки $MAPE$ та коефіцієнта Тейла U за формулами (2.22), (2.23), (2.24) відповідно. Критерієм покращення придатності побудованої моделі для прогнозування є мінімізація перерахованих вище прогнозних характеристик. Розрахунок наведених прогнозних характеристик та визначення кращих методів відбувається після натискання на кнопку «Знайти прогнозні характеристики». Назви кращих методів записуються у поле з заголовком

«Кращі методи:». Після натискання на відповідну кнопку «Очистити» відбувається очищення значень наведених прогнозних характеристик.

Таким чином, на закладці «Прогнозування часових рядів» реалізовано принцип процедурної повноти, оскільки користувач має змогу виконати прогнозування за наведеними методами, а після цього визначити кращі методи за статистичними та прогнозними характеристиками.

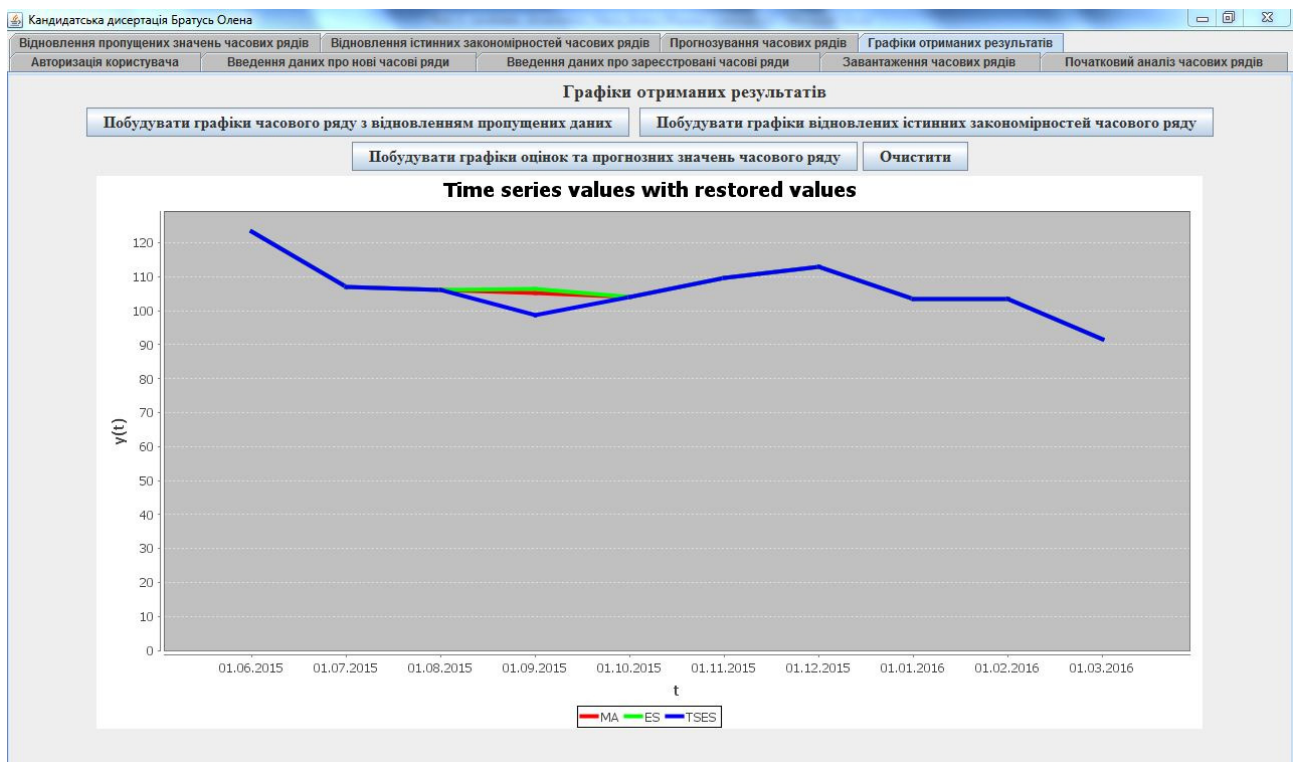


Рис. 4.17. Вигляд закладки «Графіки отриманих результатів», коли на цій закладці побудований графік часового ряду з відновленням пропущених значень

На рис. 4.17 наведено вигляд закладки «Графіки отриманих результатів». На цій закладці після натискання на кнопку «Побудувати графіки часового ряду з відновленням пропущених даних» надається можливість побудувати графіки часового ряду з відновленням пропущених даних за тими методами, за якими відбувалось відновлення на закладці «Відновлення пропущених значень часових рядів». На рис. 4.17 наведено графік часового ряду з відновленням за методами штучно пропущеним значенням часового ряду за вересень 2015 року

на основі вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU з червня 2015 року до березня 2016 року за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70]. Також на закладці «Графіки отриманих результатів» після натискання на кнопку «Побудувати графіки відновлених істинних закономірностей часового ряду» надається можливість побудувати графіки часового ряду з відновленням його істинних закономірностей за тими методами, за якими відбувалось відновлення на закладці «Відновлення істинних закономірностей часових рядів», що показано на рис. 4.18 для вибірки даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU з червня 2015 року до лютого 2016 року за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70]. На цій закладці після натискання на кнопку «Побудувати графіки оцінок та прогнозних значень часового ряду» надається можливість побудувати графіки часового ряду, графіки оцінок значень ряду та оцінок прогнозних значень ряду за тими методами, за якими відбувалось прогнозування на закладці «Прогнозування часових рядів», що показано на рис. 4.19 для вибірки даних про середньодобові ціни Лондонської біржі металів на цинк у американських доларах за тонну у робочі дні за період з 01.02.2017 по 15.02.2017 [52]. Після натискання на кнопку «Очистити» відбувається очищення закладки від того графіку, який побудований у даний момент на цій закладці.

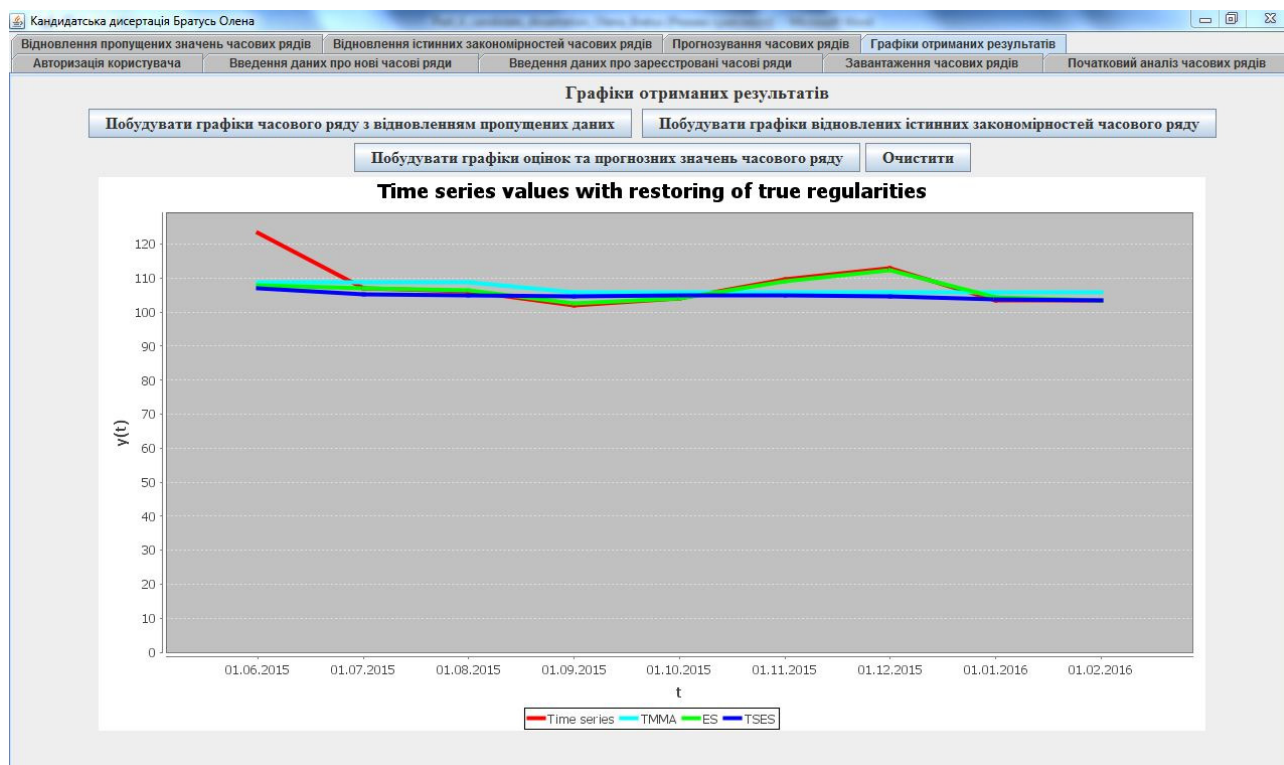


Рис. 4.18. Вигляд закладки «Графіки отриманих результатів», коли на цій закладці побудований графік часового ряду з відновленням істинних закономірностей розвитку

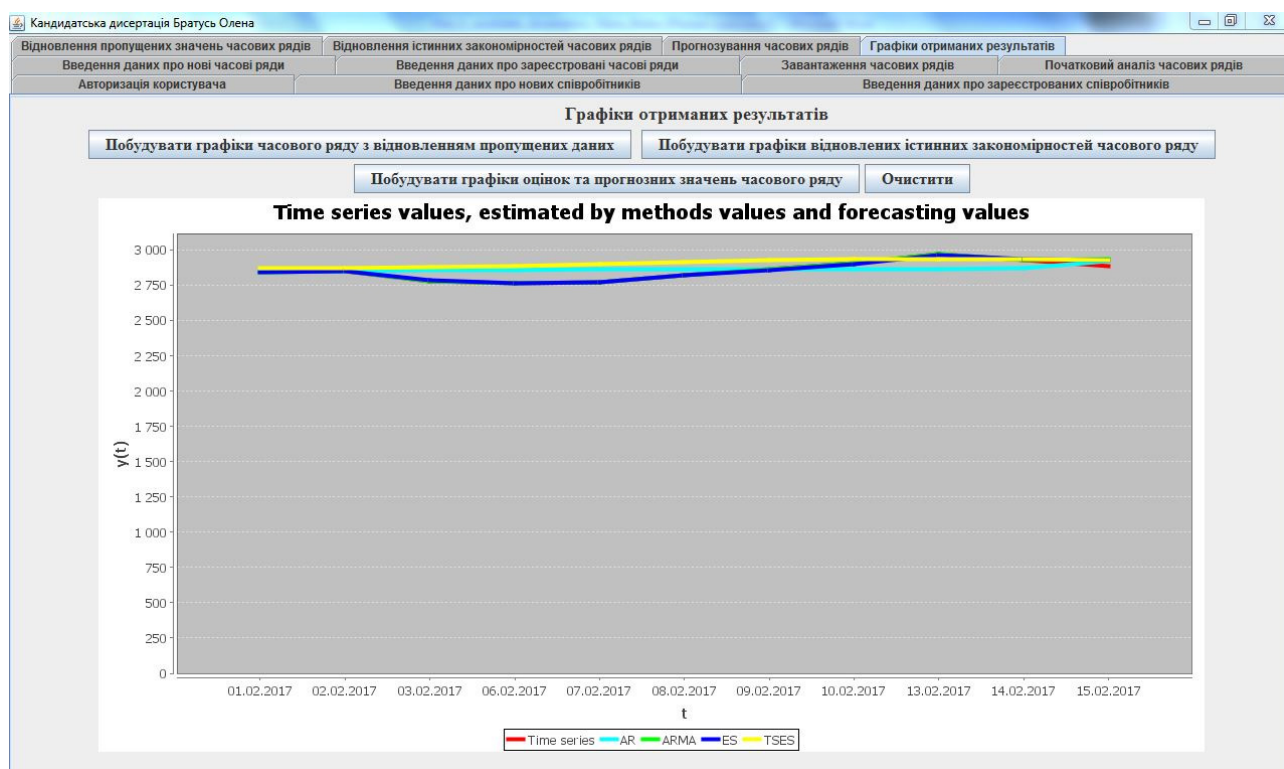


Рис. 4.19. Вигляд закладки «Графіки отриманих результатів», коли на цій закладці побудований графік часового ряду з оцінками ряду та оцінками прогнозів за методами прогнозування

Таким чином, на закладці «Графіки отриманих результатів» користувачеві надається можливість побудувати графік часового ряду з відновленням пропущених значень, графік часового ряду з відновленням істинних закономірностей розвитку, графік часового ряду з оцінками ряду та оцінками прогнозів за методами прогнозування, що дає можливість додатково візуально оцінити отримані результати.

4.4. Висновки до розділу 4

Виконано дослідження стосовно застосування системного підходу до розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів. Визначено, що для розробки такої системи підтримки прийняття рішень повинен бути виконаний попередній інформаційний аналіз. Також важливим є дотримання основних принципів

формування системної методології та її практичної реалізації під час розробки СППР. В організації, що займається аналізом та прогнозуванням часових рядів, виділено відділи інформаційних технологій, аналітики, кадрів. У відповідності до цих відділів співробітникам надаються права на аналіз та обробку даних, тобто виконується системний принцип інформаційної взаємозалежності. Створено структурну схему частини аналізу та прогнозування часових рядів у системі підтримки прийняття рішень, на якій описано застосування системного підходу до частини аналізу та прогнозування часових рядів у СППР. Описано бізнес-правила роботи з СППР та матриці подій для співробітників наведених відділів.

Виконано проектування бази даних, що є важливим етапом перед роботою над програмною реалізацією СППР. Створено ER-діаграму фізичну загальну бази даних, на якій описано таблиці та зв'язки між ними для бази даних, що є під'єднаною до розробленої системи підтримки прийняття рішень.

Створено архітектуру системи підтримки прийняття рішень, на якій показано основні етапи роботи з цією системою. Розроблена СППР містить наступні закладки: «Авторизація користувача», «Введення даних про нових співробітників», «Введення даних про зареєстрованих співробітників», «Введення даних про нові часові ряди», «Введення даних про зареєстровані часові ряди», «Завантаження часових рядів», «Початковий аналіз часових рядів», «Відновлення пропущених значень часових рядів», «Відновлення істинних закономірностей часових рядів», «Прогнозування часових рядів», «Графіки отриманих результатів». Описано права доступу до цих закладок для співробітників усіх відділів організації, що займається аналізом та прогнозуванням часових рядів, а також доступні функції при роботі з цими закладками. Наведено приклади застосування розробленої системи підтримки прийняття рішень для аналізу, відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів для даних про середньодобові ціни Лондонської біржі металів на цинк у американських доларах за тонну у робочі дні [52] та даних про місячні значення сонячних

радіофлюксів при довжині хвилі 10,7 см в одиницях виміру SFU за даними офіційного сайту Національного управління океанічних і атмосферних досліджень США [70].

Таким чином, розроблено систему підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів. Наведено докладний опис застосування системного підходу до розробки цієї системи. Наведено опис бази даних, що застосовується у цій СППР. Докладно описано структуру та функції розробленої системи підтримки прийняття рішень. У створеній СППР реалізовано можливості виконання аналізу даних, традиційні методи та розроблені нові методи відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Створена система підтримки прийняття рішень використовується у навчальному процесі та може бути використаною для роботи у реальних організаціях, що займаються аналізом та прогнозуванням часових рядів. Також розроблена система підтримки прийняття рішень реалізує цілісну системну методологію, оскільки вона задовольняє основним принципам формування сучасної системної методології та її практичної реалізації.

ВИСНОВКИ

У дисертаційній роботі розв'язано науково-прикладну задачу стосовно розробки методів оцінювання параметрів математичних моделей, що змінюються за невідомим законом, та їх застосування для підвищення якості оцінок прогнозів розвитку часових рядів шляхом удосконалення системної методології прогнозування, розробки методів відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів, розробки оригінального програмного продукту у формі системи підтримки прийняття рішень. За результатами виконаного дослідження можна зробити такі основні висновки:

1. Розроблено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два, та виконано порівняння створених методів з існуючими. Обрано кращі методи серед розроблених, та застосовано їх для розробки адаптивних фільтрів Калмана. Розроблені адаптивні фільтри Калмана показали перевагу перед традиційними методами за значеннями прогнозних характеристик під час прогнозування середньодобових цін Лондонської біржі металів на цинк ($MAPE=4,1562\%$ для адаптивного фільтра Калмана № 1; $MAPE=4,1176\%$ для адаптивного фільтра Калмана № 2).

2. Розроблено метод оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду однозалежних псевдовимірів цього параметру. Створено алгоритми прогнозування на основі фільтра Калмана з використанням розробленого методу оцінювання математичного сподівання прискорення зміни значень вибірки даних. Створені алгоритми прогнозування на основі фільтра Калмана показали перевагу перед традиційними методами за значеннями прогнозних характеристик під час прогнозування середньодобових цін на свинець за

даними Лондонської біржі металів ($MAPE=4,3465\%$ для алгоритму прогнозування № 1 на основі фільтра Калмана; $MAPE=4,2352\%$ для алгоритму прогнозування № 2 на основі фільтра Калмана).

3. Розроблено метод двобічного експоненційного згладжування. Створено інтегральний критерій оцінки адекватності моделі. Описано застосування створеного методу для відновлення істинних закономірностей та прогнозування часових рядів. Розроблено алгоритм для відновлення пропущених значень часового ряду з використанням методу двобічного експоненційного згладжування. Метод двобічного експоненційного згладжування за значеннями статистичних характеристик показав перевагу перед традиційними методами під час відновлення середньодобових значень цін на цинк за даними Лондонської біржі металів.

4. Розроблено метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Створений метод показав перевагу перед традиційними методами за значеннями статистичних характеристик під час відновлення значень взаємозалежних показників сталого розвитку за даними Світового центру даних. Розроблено метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Створений метод показав перевагу перед методом експоненційного згладжування за значеннями прогнозних характеристик під час прогнозування показників сталого розвитку.

5. Розроблено метод ковзного двобічного експоненційного згладжування та алгоритм для відновлення пропущених значень ряду з використанням цього методу. Створений алгоритм показав перевагу перед методом експоненційного згладжування за значеннями статистичних характеристик під час відновлення даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см за даними Національного управління океанічних і атмосферних досліджень Сполучених Штатів Америки. Розроблено критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів. Метод

ковзного двобічного експоненційного згладжування показав перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей індексу F10.7 за значенням критерію близькості. За значеннями прогнозних характеристик розроблений метод показав перевагу перед традиційним експоненційним підходом при прогнозуванні значень індексу F10.7.

6. Розроблено метод адаптивного ковзного двобічного експоненційного згладжування та алгоритм для відновлення пропущених значень часових рядів з використанням цього методу. За значенням критерію близькості створений метод показав перевагу перед методом 13-місячного ковзного середнього під час відновлення істинних закономірностей чисел Вольфа та даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см. За значеннями статистичних характеристик створений алгоритм з використанням методу адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед методом експоненційного згладжування при відновленні пропущених значень індексу F10.7. За значеннями прогнозних характеристик метод адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед традиційним експоненційним підходом при прогнозуванні місячних значень сонячних радіофлюксів при довжині хвилі 10,7 см.

7. Визначено принципи системної методології, які застосовані під час розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів. Розроблено архітектуру і структурну схему частини аналізу та прогнозування часових рядів цієї системи підтримки прийняття рішень, створено різні рівні доступу до системи підтримки прийняття рішень в залежності від відділів для співробітників цієї організації. Виконано її програмну реалізацію мовою програмування Java, та створено опис для її користувачів. У створеній системі підтримки прийняття рішень реалізовано можливості виконання аналізу даних, традиційні методи та розроблені нові методи відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Наведено приклади

її застосування для аналізу, відновлення пропущених значень, відновлення істинних закономірностей, прогнозування фінансово-економічних та сонячних даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Братусь О.В., Подладчиков В.М. Побудова багатовимірної моделі на основі фільтра Калмана й аналіз алгоритмів оцінювання її параметрів. *Наукові вісті НТУУ «КПІ»*. 2013. № 5 (91). С. 28–34.
2. Братусь О.В., Подладчиков В.М. Ідентифікація змінних параметрів моделі для побудови алгоритму прогнозування. *Системні дослідження та інформаційні технології*. 2015. № 3. С. 131–141.
3. Братусь О.В., Подладчиков В.М., Бідюк П.І. Метод двобічного експоненційного згладжування для відновлення динамічних процесів. *Наукові вісті НТУУ «КПІ»*. 2016. № 6 (110). С. 15–21.
4. Братусь О.В. Система підтримки прийняття рішень з адаптивними блоками відновлення та прогнозування сонячних радіофлюксів. *Радіoeлектроніка, інформатика, управління*. 2017. № 3. С. 36–43.
5. Братусь О.В., Бідюк П.І., Болдак А.О. Розробка методів відновлення пропущених значень і прогнозування для взаємозалежних часових рядів. *Міжнародний науково-технічний журнал «Проблеми керування та інформатики»*. 2017. № 5. С. 13–21.
6. Bratus O. Development of adaptive moving two-sided exponential smoothing method for restoring and forecasting of time series. *EUREKA: Physics and Engineering*. 2017. No. 5. P. 13–21.
7. Братусь О.В., Подладчиков В.М. Розробка методів прогнозування на основі фільтра Калмана. *Системний аналіз та інформаційні технології*: матеріали 15-ї Міжнародної науково-технічної конференції SAIT 2013, м. Київ, ННК «ІПСА» НТУУ «КПІ», 27-31 травня 2013 р. Київ: ННК «ІПСА» НТУУ «КПІ», 2013. С. 251.
8. Братусь О.В., Подладчиков В.М. Пошук закономірностей динамічних процесів на основі двобічного експоненційного згладжування. *Контроль і управління в складних системах (КУСС-2014)*: тези доповідей XII міжнародної конференції, м. Вінниця, 14-16 жовтня 2014 р. Вінниця: ВНТУ, 2014. С. 28.

9. Братусь О.В. Система підтримки прийняття рішень для аналізу та прогнозування цін на метали. *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту*: матеріали міжнародної наукової конференції, с. Залізний Порт, 24-28 травня 2016 р. Херсон: ПП Вишемирський В.С., 2016. С. 185–186.
10. Братусь О.В. Ідентифікація змінної дисперсії шуму вимірів моделі у задачах прогнозування. *Математика. Інформаційні технології. Освіта*: тези доповідей V міжнародної конференції, м. Луцьк - с. Світязь, 5-7 червня 2016 р. Луцьк: ПП Іванюк В.П., 2016. С. 58–59.
11. Братусь О.В., Подладчиков В.М., Подладчикова Т.В. Метод ковзного двобічного експоненційного згладжування для відновлення закономірностей динамічних процесів. *Математичне та імітаційне моделювання систем. МОДС 2016*: тези доповідей одинадцятої міжнародної науково-практичної конференції, с. Жукин, 27 червня - 1 липня 2016 р. Чернігів: ЧНТУ, 2016. С. 28–31.
12. Згуровський М.З., Панкратова Н.Д. Основи системного аналізу. Київ: Видавнича група ВНУ, 2007. 544 с.
13. Сорока К.О. Основи теорії систем і системного аналізу. Харків: ХНАМГ, 2004. 291 с.
14. Клир Дж. Системология. Автоматизация решения системных задач. Москва: Радио и связь, 1990. 540 с.
15. Бідюк П.І., Савенко О.І., Баклан В.І. Часові ряди: моделювання та прогнозування. Київ: ЕКМО, 2004. 114 с.
16. Стаття «Тренд (статистика)». URL:
[https://uk.wikipedia.org/wiki/%D0%A2%D1%80%D0%B5%D0%BD%D0%B4_\(%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0\)](https://uk.wikipedia.org/wiki/%D0%A2%D1%80%D0%B5%D0%BD%D0%B4_(%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0))
17. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. 2. Москва: Финансы и статистика, 1986. 366 с.

18. Бідюк П.І. Системний підхід до прогнозування на основі моделей часових рядів. *Системні дослідження та інформаційні технології*. 2003. № 3. С. 88–110.
19. Chatfield C. *The Analysis of Time Series: An Introduction*, Sixth Edition. Boca Raton, Florida: Chapman & Hall/CRC Press, 2003. 333 p.
20. Enders W. *Applied econometric time series*. New York: John Wiley & Sons, Inc., 1995. 434 p.
21. Restrepo J.M., Venkataramani S., Comeau D., Flaschka H. Defining a trend for time series using the intrinsic time-scale decomposition. *New Journal of Physics*. 2014. Vol. 16, No. 8. P. 1–28. DOI: <https://doi.org/10.1088/1367-2630/16/8/085004>
URL: <http://iopscience.iop.org/article/10.1088/1367-2630/16/8/085004>
22. Caivano M., Harvey A., Luati A. Robust time series models with trend and seasonal components. *SERIEs*. 2016. Vol. 7, Issue 1. P. 99–120.
23. Tsay R.S. *Analysis of Financial Time Series*. New York: John Wiley & Sons, Inc., 2010. 715 p.
24. Бідюк П.І., Романенко В.Д., Тимошук О.Л. Аналіз часових рядів. Київ: НТУУ «КПІ», 2013. 600 с.
25. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. Москва: Финансы и статистика, 2003. 416 с.
26. Бідюк П.І., Коршевнік Л.О. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень. Київ: ННК «ІПСА» НТУУ «КПІ», 2010. 340 с.
27. Hathaway D.H., Wilson R.M., Reichmann E.J. A synthesis of solar cycle prediction techniques. *Journal of Geophysical Research*. 1999. Vol. 104, No. A10. P. 22375–22388. DOI: <https://doi.org/10.1029/1999JA900313>
28. Hathaway D.H., Wilson R.M., Reichmann E.J. The shape of the sunspot cycle. *Solar Physics*. 1994. Vol. 151, No. 1. P. 177–190. DOI: <https://doi.org/10.1007/BF00654090>
29. Seber G.A.F., Wild C.J. *Nonlinear Regression*. Hoboken: John Wiley & Sons, Inc., 2003. 768 p.

30. Brown R.G. Exponential Smoothing for Predicting Demand. Cambridge, Massachusetts: Arthur D. Little Inc., 1956. 15 p.
31. Holt C.C. Forecasting Trends and Seasonal by Exponentially Weighted Averages. *International Journal of Forecasting*. 1957. Vol. 20, No. 1. P. 5–10.
32. Gardner E.S., Jr. Exponential Smoothing: the State of the Art. *Journal of Forecasting*. 1985. Vol. 4, No. 1. P. 1–28.
33. Gardner E.S., Jr. Exponential smoothing: The state of the art - Part II. *Journal of Forecasting*. 2006. Vol. 22, No. 4. P. 637–666.
34. Згуровский М.З., Подладчиков В.Н. Аналитические методы калмановской фильтрации для систем с априорной неопределённостью. Киев: Наукова думка, 1995. 283 с.
35. Kalman R.E. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*. 1960. Vol. 82D, March. P. 34–45.
36. Ивахненко А.Г. Непрерывность и дискретность. Киев: Наукова думка, 1990. 223 с.
37. Ивахненко А.Г., Юрачковский Ю.П. Моделирование систем по экспериментальным данным. Москва: Радио и связь, 1987. 120 с.
38. Ивахненко А.Г., Мюллер И.А. Самоорганизация прогнозирующих моделей. Киев: Техника, 1985. 223 с.
39. Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д. Принятие решений на основе самоорганизации. Москва: Советское радио, 1976. 280 с.
40. Зайченко Ю.П. Основи проектування інтелектуальних систем. Київ: Слово, 2004. 352 с.
41. Степашко В.С. Структурна ідентифікація моделей як задача відновлення сигналу в умовах неповноти інформації. *Наукові праці ДонНТУ. Серія: Обчислювальна техніка та автоматизація*. Донецьк: ДонНТУ, 2002. Вип. 48. С. 345–353.
42. Степашко В.С. Теоретические аспекты МГУА как метода индуктивного моделирования. *Управляющие системы и машины*. 2003. № 2. С. 31–44.

43. Бідюк П.І., Коршевніук Л.О., Гожий О.П. Комп'ютерні системи підтримки прийняття рішень. Київ: ННК «ІПСА» НТУУ «КПІ», 2011. 340 с.
44. Бідюк П.І., Демківський Є.О. Система підтримки прийняття рішень для прогнозування нестационарних процесів. *Наукові праці Чорноморського державного університету імені Петра Могили. Серія: Комп'ютерні технології*. 2008. Т. 90. Вип. 77. С. 137–159.
45. Берзлев О.Ю., Білощицький А.О. Інформаційна система для прогнозування і прийняття рішень у фінансовій сфері. *Управління розвитком складних систем*. 2014. Вип. 18. С. 106–111.
46. Тоценко В.Г. Методи та засоби підтримки прийняття рішень. *Реєстрація, зберігання і обробка даних*. 2007. Т. 9. № 3. С. 98–104.
47. Демиденко М.А. Системи підтримки прийняття рішень. Дніпропетровськ: НГУ, 2016. 104 с.
48. Фильтрация и стохастическое управление в динамических системах: под ред. К.Т. Леондеса. Москва: Мир, 1980. 407 с.
49. Мехра Р. Идентификация и адаптивная фильтрация Калмана. *Механика (сб. переводов статей)*. 1971. № 3. С. 34–51.
50. Oussalah M., De Schutter J. Adaptive Kalman Filter for Noise Identification. *International Conference on Noise and Vibration Engineering: proceedings*, 2000. P. 1225–1232.
51. Almagbile A., Wang J., Ding W. Evaluating the Performances of Adaptive Kalman Filter Methods in GPS/INS Integration. *Journal of Global Positioning Systems*. 2010. Vol. 9, No. 1. P. 33–40.
52. Офіційний сайт Лондонської біржі металів. Розділ «Місячна середня розрахункова ціна».
URL: <https://www.lme.com/Market-Data/Reports-and-data/Monthly-averages>
53. Astrom K.J., Eykhoff P. System identification – a survey. *Automatica*. 1971. Vol. 7, P. 123–162.

54. Anderson W.N., Jr., Kleindorfer G.B., Kleindorfer P.R., Woodroffe M.B. Consistent Estimates of the Parameters of a Linear System. *The Annals of Mathematical Statistics*. 1969. Vol. 40, No. 6. P. 2064–2075.
55. Podladchikova T., Van der Linden R. A Kalman Filter Technique for Improving Medium-Term Predictions of the Sunspot Number. *Solar Physics*. 2012. Vol. 277, P. 397–416.
56. Бендат Дж., Пирсол А. Прикладной анализ случайных данных: пер. с англ. д-ра физ.-мат. наук В.Е. Привальского и А.И. Кочубинского под ред. акад. И.Н. Коваленко. Москва: Мир, 1989. 540 с.
57. Behrouz F.-B. Adaptive Filters. New York: John Wiley & Sons, Inc., 2013. 802 p.
58. Hyndman R.J., Koehler A.B., Ord J.K., Snyder R.D. Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting*. 2005. Vol. 24, No. 1. P. 17–37.
59. Kolassa S. Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*. 2011. Vol. 27, No. 2. P. 238–251.
60. Pankratova N.D., Podladchikova T.V., Strelkov D.G. Quasioptimal smoothing as a tool for the analysis of complex semistructured dynamic processes. *Cybernetics and Systems Analysis*. 2009. Vol. 45, No. 6. P. 916–923.
61. Mickaityte A., Zavadskas E.K., Kaklauskas A., Tupenaite L. The concept model of sustainable buildings refurbishment. *International Journal of Strategic Property Management*. 2008. Vol. 12, Issue 1. P. 53–68.
62. Prior M., Farmer J., Godden D.J., Taylor J. More than health: the added value of health services in remote Scotland and Australia. *Health Place*. 2010. Vol. 16, Issue 6. P. 1136–1144.
63. Oven K.J., Curtis S.E., Reaney S., Riva M. et al. Climate change and health and social care: Defining future hazard, vulnerability and risk for infrastructure systems supporting older people's healthcare in England. *Applied Geography*. 2012. Vol. 33, Issue 1. P. 16–24.

64. Kristensen T., Olsen K.R., Kilsmark J., Pedersen K.M. Economies of scale and optimal size of hospitals: Empirical results for Danish public hospitals. *Health Economics Papers*. 2008. Vol. 3. 31 p.
65. Rees P., Van der Gaag N., De Beer J., Heins F. European Regional Populations: Current Trends, Future Pathways, and Policy Options. *European Journal of Population*. 2012. Vol. 4, Issue 4. P. 385-416.
66. Сайт Світового центру даних. URL: <http://wdc.org.ua/>.
67. Salvador R.J. A mathematical model of the sunspot cycle for the past 1000 yr. *Pattern recognition in physics*. 2013. Vol. 1, No. 1. P. 117–122. DOI: <https://doi.org/10.5194/prp-1-117-2013>
68. Киян М.А., Фабричева Е.В., Подладчиков В.Н. Применения алгоритмов сглаживания для предварительной обработки статистических солнечных данных. *Системні дослідження та інформаційні технології*. 2014. № 2. С. 68–76.
69. Noble P.L., Wheatland M.S. A Bayesian Approach to Forecasting Solar Cycles Using a Fokker–Planck Equation. *Solar Physics*. 2012. Vol. 276, No. 1. P. 363–381. DOI: <https://doi.org/10.1007/s11207-011-9884-5>
70. Офіційний сайт Національного управління океанічних і атмосферних досліджень США. URL: <ftp://ftp.swpc.noaa.gov/pub/weekly/RecentIndices.txt>
71. Hurst J.M. The Profit Magic of Stock Transaction Timing. Englewood Cliffs: Prentice-Hall, Inc., 1970. 224 p.
72. Jouve L., Brun A.S., Arlt R., Brandenburg A. et al. A solar mean field dynamo benchmark. *Astronomy & Astrophysics*. 2008. Vol. 483. P. 949–960.
73. Seehafer N., Pipin V.V. An advective solar-type dynamo without the α effect. *Astronomy & Astrophysics*. 2009. Vol. 508. P. 9–16.
74. Schatten K.H. Modelling a Shallow Solar Dynamo. *Solar Physics*. 2009. Vol. 255, P. 3–38.
75. Dikpati M., Gilman P.A., De Toma G. The Waldmier Effect: An Artifact of the Definition of Wolf Sunspot Number? *The Astrophysical Journal Letters*. 2008. Vol. 673, No. 1. L99–L101. DOI: <https://doi.org/10.1086/527360>

76. Clette F., Berghmans D., Vanlommel P., Van der Linden R.A.M. et al. From the Wolf number to the International Sunspot Index: 25 years of SIDC. 2007. *Advances in Space Research*. Vol. 40, No. 7. P. 919–928. DOI: <https://doi.org/10.1016/j.asr.2006.12.045>
77. Usoskin I.G., Mursula K., Arlt R., Kovaltsov G.A. A Solar Cycle Lost in 1793-1800: Early Sunspot Observations Resolve the Old Mystery. *The Astrophysical Journal Letters*. 2009. Vol. 700. L154–L157. DOI: <https://doi.org/10.1088/0004-637X/700/2/L154>
78. Clette F., Svalgaard L., Vaquero J.M., Cliver E.W. Revisiting the Sunspot Number. A 400-year perspective on the solar cycle. *Space Science Reviews*. 2014. Vol. 186, Issue 1–4. P. 35–103. DOI: <https://doi.org/10.1007/s11214-014-0074-2>
79. Clette F., Lefevre L., Cagnotti M., Cortessi S., Bulling A. The revised Brussels-Locarno Sunspot Number (1981–2015). *Solar Physics*. 2016. Vol. 291, Issue 9–10. P. 2733–2761. DOI: <https://doi.org/10.1007/s11207-016-0875-4>
80. Clette F., Lefevre L. The New Sunspot Number: Assembling All Corrections. *Solar Physics*. 2016. Vol. 291, Issue 9–10. P. 2629–2651. DOI: <https://doi.org/10.1007/s11207-016-1014-y>
81. De Wit T.D., Lefevre L., Clette F. Uncertainties in the Sunspot Numbers: Estimation and implications. *Solar Physics*. 2016. Vol. 291, Issue 9–10. P. 2709–2731. DOI: <https://doi.org/10.1007/s11207-016-0970-6>
82. Lockwood M., Owens M.J., Barnard L. Centennial variations in Sunspot Number, Open Solar Flux, and streamer belt width: 1. Correction of the Sunspot Number record since 1874. *Journal of Geophysical Research: Space Physics*. 2014. Vol. 119, No. 7. P. 5172–5182. DOI: <https://doi.org/10.1002/2014JA019970>
83. Lockwood M., Scott C.J., Owens M.J., Barnard L. et al. Tests of Sunspot Number Sequences: 1. Using Ionosonde Data. *Solar Physics*. 2016. Vol. 291, Issue 9–10. P. 2785–2809. DOI: <https://doi.org/10.1007/s11207-016-0855-8>
84. Dikpati M., De Toma G., Gilman P.A. Polar flux, cross-equatorial flux, and dynamo generated tachocline toroidal flux as predictors of solar cycles. *Astrophysical Journal*. 2008. Vol. 675, No. 1. P. 920–930.

85. Yeates A.R., Nandy D., Mackay D.H. Exploring the Physical Basis of Solar Cycle Predictions: Flux Transport Dynamics and Persistence of Memory in Advection-versus Diffusion-dominated Solar Convection Zones. *Astrophysical Journal*. 2008. Vol. 673, No. 1. P. 544–556. DOI: <https://doi.org/10.1086/524352>
86. Johnson R.W. Power law relating 10.7 cm flux to sunspot number. *Astrophysics and Space Science*. 2011. Vol. 332. P. 73–79.
DOI: <https://doi.org/10.1007/s10509-010-0500-1>
87. Pesnell W.D. Predictions of Solar Cycle 24: How are we doing? *Space Weather*. 2016. Vol. 14, Issue 1. P. 10–21.
88. Эккель Б. Философия Java. Библиотека программиста. 4-е изд. Санкт-Петербург: Питер, 2012. 640 с.
89. Блинов И.Н., Романчик В.С. Java. Промышленное программирование: практ. пособие. Минск: УниверсалПресс, 2007. 704 с.

Додаток А

**Акт впровадження результатів дисертаційної роботи у
Навчально-науковому комплексі «Світовий центр даних з геоінформатики
та сталого розвитку»**



13.03.2018 № 0301/18

Акт

впровадження результатів дисертаційної роботи

Братусь Олени Вікторівни,

представленої на здобуття наукового ступеня кандидата технічних наук

Дисертаційна робота Братусь О. В. присвячена розробці методів відновлення пропущених значень та прогнозування нестаціонарних часових рядів. У процесі виконання дисертаційних досліджень створено метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Виконано практичне застосування цього методу для відновлення штучно створених пропусків та фактично пропущених значень взаємозалежних показників сталого розвитку Хорватії (багатовимірні часові ряди).

У процесі виконання дисертаційних досліджень також розроблено метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. З використанням цього методу виконано прогнозування значень взаємозалежних показників сталого розвитку Хорватії.

Створені методи апробовано на практиці під час виконання аналізу та прогнозування часових рядів у ННК «Світовий центр даних з геоінформатики та сталого розвитку», отримано високоякісні результати відновлення пропущених значень та оцінок прогнозів для взаємозалежних показників сталого розвитку Хорватії. ННК «СЦД-Україна» підтверджує доцільність практичного застосування та ефективність створених методів, оскільки вони дають можливість отримувати точніші оцінки пропущених значень та прогнозів часових рядів порівняно з традиційними методами.

Директор



К. В. Єфремов

Додаток Б

Довідка про впровадження результатів дисертаційної роботи у навчальний процес Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ * МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ



**ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ім. Ігоря Сікорського»**

№ 48/4040
«22» 03 2018



Затверджую
Заступник Директора ІПСА
НТУУ «КПІ ім. Ігоря Сікорського»
з навчально-наукової роботи,
д.т.н., професор

В. Д. Романенко

ДОВІДКА

**про впровадження результатів дисертаційної роботи
Братусь Олени Вікторівни
у навчальний процес ІПСА НТУУ «КПІ ім. Ігоря Сікорського»**

У процесі виконання дисертаційної роботи Братусь Олена Вікторівна отримала ряд нових наукових результатів стосовно відновлення пропущених значень вибірок даних, відновлення істинних закономірностей розвитку та прогнозування часових рядів. Отримані результати впроваджено у навчальний процес кафедри Математичних методів системного аналізу Інституту прикладного системного аналізу НТУУ «КПІ ім. Ігоря Сікорського».

Зокрема, запропоновано методи відновлення пропущених значень, відновлення істинних закономірностей розвитку та прогнозування часових рядів. Порівняно з традиційними методами запропоновано враховувати інформацію про часовий ряд, отриману у зворотному часі за допомогою методів двобічного експоненційного згладжування, ковзного двобічного експоненційного згладжування та адаптивного ковзного двобічного експоненційного згладжування. Також побудовано адаптивні оптимальні фільтри Калмана для короткострокового прогнозування фінансових часових рядів.

На основі запропонованих методів і математичних моделей розроблено лабораторні роботи для таких дисциплін з навчального плану Інституту прикладного системного аналізу: «Аналіз часових рядів», «Прогнозування у фінансово-економічній сфері» та «Аналіз нелінійних часових рядів». Розроблене програмне і алгоритмічне забезпечення використано для створення сучасної інформаційної технології аналізу фінансових даних, яка взята за основу для лабораторних та курсових робіт з навчальної дисципліни «Системи і методи підтримки прийняття рішень».

Професор ІПСА
НТУУ «КПІ ім. Ігоря Сікорського»,
д.т.н.

В. Я. Данилов

Професор ІПСА
НТУУ «КПІ ім. Ігоря Сікорського»,
д.т.н.

П. І. Бідок

Додаток В

Список публікацій здобувача та апробація результатів дисертації

Список публікацій здобувача

1. Братусь О.В., Подладчиков В.М. Побудова багатовимірної моделі на основі фільтра Калмана й аналіз алгоритмів оцінювання її параметрів. *Наукові вісті НТУУ «КПІ»*. 2013. № 5 (91). С. 28–34.
2. Братусь О.В., Подладчиков В.М. Ідентифікація змінних параметрів моделі для побудови алгоритму прогнозування. *Системні дослідження та інформаційні технології*. 2015. № 3. С. 131–141.
3. Братусь О.В., Подладчиков В.М., Бідюк П.І. Метод двобічного експоненційного згладжування для відновлення динамічних процесів. *Наукові вісті НТУУ «КПІ»*. 2016. № 6 (110). С. 15–21.
4. Братусь О.В. Система підтримки прийняття рішень з адаптивними блоками відновлення та прогнозування сонячних радіофлюксів. *Радіoeлектроніка, інформатика, управління*. 2017. № 3. С. 36–43.
5. Братусь О.В., Бідюк П.І., Болдак А.О. Розробка методів відновлення пропущених значень і прогнозування для взаємозалежних часових рядів. *Міжнародний науково-технічний журнал «Проблеми керування та інформатики»*. 2017. № 5. С. 13–21.
6. Bratus O. Development of adaptive moving two-sided exponential smoothing method for restoring and forecasting of time series. *EUREKA: Physics and Engineering*. 2017. No. 5. P. 13–21.
7. Братусь О.В., Подладчиков В.М. Розробка методів прогнозування на основі фільтра Калмана. *Системний аналіз та інформаційні технології*: матеріали 15-ї Міжнародної науково-технічної конференції SAIT 2013, м. Київ, ННК «ІПСА» НТУУ «КПІ», 27-31 травня 2013 р. Київ: ННК «ІПСА» НТУУ «КПІ», 2013. С. 251.

8. Братусь О.В., Подладчиков В.М. Пошук закономірностей динамічних процесів на основі двобічного експоненційного згладжування. *Контроль і управління в складних системах (КУСС-2014)*: тези доповідей XII міжнародної конференції, м. Вінниця, 14-16 жовтня 2014 р. Вінниця: ВНТУ, 2014. С. 28.
9. Братусь О.В. Система підтримки прийняття рішень для аналізу та прогнозування цін на метали. *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту*: матеріали міжнародної наукової конференції, с. Залізний Порт, 24-28 травня 2016 р. Херсон: ПП Вишемирський В.С., 2016. С. 185–186.
10. Братусь О.В. Ідентифікація змінної дисперсії шуму вимірів моделі у задачах прогнозування. *Математика. Інформаційні технології. Освіта*: тези доповідей V міжнародної конференції, м. Луцьк - с. Світязь, 5-7 червня 2016 р. Луцьк: ПП Іванюк В.П., 2016. С. 58–59.
11. Братусь О.В., Подладчиков В.М., Подладчикова Т.В. Метод ковзного двобічного експоненційного згладжування для відновлення закономірностей динамічних процесів. *Математичне та імітаційне моделювання систем. МОДС 2016*: тези доповідей одинадцятої міжнародної науково-практичної конференції, с. Жукин, 27 червня - 1 липня 2016 р. Чернігів: ЧНТУ, 2016. С. 28–31.

Апробація результатів дисертації

Основні положення дисертаційної роботи були представлені на наукових конференціях та семінарах:

- 15^й міжнародній науково-технічній конференції SAIT 2013 (Україна, м. Київ, 27-31 травня 2013 р.);
- XII міжнародній конференції «Контроль і управління в складних системах (КУСС-2014)» (Україна, м. Вінниця, 14-16 жовтня 2014 р.);

- міжнародній науковій конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (Україна, с. Залізний Порт, 24-28 травня 2016 р.);
- V^й міжнародній конференції «Математика. Інформаційні технології. Освіта» (Україна, м. Луцьк - с. Світязь, 5-7 червня 2016 р.);
- Одинадцятій міжнародній науково-практичній конференції «Математичне та імітаційне моделювання систем. МОДС 2016» (Україна, с. Жукин, 27 червня - 1 липня 2016 р.);
- науковому семінарі «Системні дослідження та інформаційні технології» в Інституті прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (Україна, м. Київ, 21 березня 2018 р.).